

***Probe-Signal Investigation of an Attentional Filter for  
Fundamental Frequency***

---

Rachel N. Peddie

A thesis submitted in partial fulfilment of the requirements for the degree of  
**Master of Audiology**

For the Department of Communication Disorders  
University of Canterbury  
2014

## Acknowledgements

---

This thesis would not have been possible without the interest and support of others. Firstly I would like to thank my primary supervisor, Dr Donal Sinex, for his expert knowledge, patience and for the many hours he spent answering my questions and providing feedback on drafts throughout the thesis process. Secondly, I would like to thank my secondary supervisor, Dr Greg O’Beirne, for his time spent proof-reading and providing feedback. Without their help, and the help of the participants, this study would not have been possible. I would also like to acknowledge Dr Emily Lin and Elena Moltchanova for their advice and support in conducting the statistical analyses of my findings.

Sincere thanks goes to my wonderful friends, family and classmates, for their amazing support throughout the past two years of this Master’s degree. Finally, a special thank you goes out to my partner, Jason, for his unconditional love and encouragement throughout the ups and downs of my thesis journey.

## Abstract

---

When listeners are detecting tones of a given frequency in noise, they operate with a narrowband attentional filter which is tuned to the frequency of the attended tone. This means that tones with frequencies which match that which is being attended to, will be detected, whereas tones with frequencies outside the filter will be detected at chance levels. In the current study, attentional filters were measured on two auditory dimensions. The first experiment employed a modified version of the methods of previous studies to measure attentional filters in the frequency dimension (Greenberg & Larkin, 1968). The mean results replicated those of previous experiments, revealing an attentional filter for frequency. In the second experiment, the methods of the first experiment were used to investigate whether there is an attentional filter in the fundamental frequency ( $f_0$ ) dimension. To test this, sentences with a given  $f_0$  were presented in noise. Infrequently, sentences were presented with  $f_0$ s which the listeners were not attending to. The changes in  $f_0$  ( $\Delta f_0$ ) tested were  $\pm 5$  Hz and  $\pm 10$  Hz, with respect to the expected 220 Hz  $f_0$ . The effect of  $\Delta f_0$  was investigated using sentence identification scores. Mean results indicated that listeners were best at identifying the sentences with an expected  $f_0$ , and sentences with  $\Delta f_0$ s of 10 Hz from the expected  $f_0$ . Sentences with  $\Delta f_0$ s of 5 Hz from the expected  $f_0$ , on the other hand, were more poorly detected. This could be due to the presence of an attentional filter for  $f_0$ , which has a narrow bandwidth, ranging between 10 and 20 Hz. An attentional filter for  $f_0$  may have different properties than those for attentional filters for frequency. The complex nature of sentence materials might change the way that attention is allocated across the  $f_0$  distribution, so that greater changes in  $f_0$  may be enough to switch the listeners' attention to that  $f_0$ . Overall, the results did not approximate the typical attentional filter shape which was found for frequency. Therefore, the results cannot be used to demonstrate the existence of attentional filters for  $f_0$ . The presence of an attentional filter for  $f_0$ , however, cannot be ruled out. Further research is

needed, using a greater variety of  $\Delta f_0$ s to confirm and further investigate the presence, and properties of an attentional filter for  $f_0$ .

## Table of Contents

|   |     |
|---|-----|
| Abstract .....  | i   |
| List of Figures .....   | v   |
| List of Tables .....  | vii |
| Chapter One: Introduction .....   | 1   |
| 1.1 Frequency Selectivity, the Critical Band and the Auditory Filter .....      | 2   |
| 1.2 Attention.....  | 4   |
| 1.2.1 Theories of Auditory Attention.....                                       | 5   |
| 1.3 Auditory Stream Segregation .....   | 8   |
| 1.3.1. Cherry’s Description of the “Cocktail Party Effect” .....                | 8   |
| 1.3.2. Bottom-Up versus Top-Down Segregation .....                              | 11  |
| 1.4 Fundamental Frequency .....   | 12  |
| 1.4.1 The Role of f0 on the Ability to Identify Speech in Noise .....           | 12  |
| 1.5 Detection of Tones in Noise.....  | 17  |
| 1.5.1 The Use of a Probe-Signal Method to Estimate the Attentional Filter ..... | 18  |
| 1.6 Hypotheses and Rationale .....  | 37  |
| Chapter Two: Methods .....  | 39  |
| 2.0.1 Participants .....  | 39  |
| 2.0.2 Instrumentation.....  | 40  |
| 2. 1. Experiment 1.....   | 40  |
| 2.1.1 Stimuli.....  | 40  |
| 2.1.2 Procedure .....   | 41  |
| 2.2 Experiment 2 .....  | 43  |
| 2.2.1 Stimuli.....  | 43  |
| 2.2.2 Signal Manipulation.....  | 44  |
| 2.2.3 Procedure .....   | 45  |
| 2.3 Data Analysis .....   | 47  |
| Chapter Three: Results.....   | 48  |
| 3.1 Experiment 1 .....  | 48  |
| 3.1.2 Statistical Analysis for Experiment 1 .....                               | 54  |
| 3.2 Experiment 2 .....  | 56  |
| 3.2.1 Statistical Analysis for Experiment 2 .....                               | 61  |
| Chapter Four: Discussion.....   | 64  |
| 4.1 Experiment 1 .....  | 64  |
| 4.1.1 Experimental Findings .....   | 64  |
| 4.1.2 Selective Attention.....  | 69  |
| 4.1.3 Individual Differences .....  | 70  |

|       |   |     |
|-------|---|-----|
| 4.1.4 | Conclusion .....                                    | 71  |
| 4.2   | Experiment 2 .....                                  | 72  |
| 4.2.1 | Experimental Findings .....                         | 72  |
| 4.2.2 | Attention .....                                     | 73  |
| 4.2.3 | Fundamental Frequency .....                         | 75  |
| 4.2.4 | Possible Contributing Factors .....                 | 76  |
| 4.2.5 | Conclusions and Directions for Future Research..... | 79  |
| 4.3   | Summary and Conclusions.....                        | 80  |
|       | References.....                                     | 82  |
|       | Appendix A. Participant Information Sheet.....      | I   |
|       | Appendix B. Participant Consent Form .....          | III |
|       | Appendix C. Instructions to Participants .....      | IV  |
|       | Appendix D. Summary of Speech Manipulation .....    | V   |

## List of Figures

|  |    |
|--|----|
| <b>Figure 1.</b> Characteristics of a low-pass filter (A); a band-pass filter (B); a high-pass filter (C).....   | 3  |
| <b>Figure 2.</b> The results of the probe-signal method for one subject's performance (percent correct) using 14 probe signal frequencies and a target signal of 1100 Hz (Fig 2 from Greenberg and Larkin, 1968).....  | 19 |
| <b>Figure 3.</b> Mean percentage of trials on which the subjects correctly identified the interval that contained the signal. Detection hovers about 90% for expected signals across the frequencies; however detection rates fall away when the signals are unexpected probe.....   | 20 |
| <b>Figure 4.</b> Listening bands/psychometric functions for a 1 kHz target (closed symbols) and probe tones (open symbols) derived using a probe-signal method for three participants (Fig. 4 from Dai et al., 1991).....  | 23 |
| <b>Figure 5.</b> Example of attention functions plotted with attenuation (dB) as a function of frequency (Fig 4 from Schlauch & Hafter, 1991).....   | 24 |
| <b>Figure 6.</b> The bandwidth of the attentional filter (labelled attention band) derived from attenuation and plotted as a function of target frequency. Plotted for comparison are measures of the critical bandwidth (Zwicker & Terhardt, 1982) and auditory filter (Patterson & Moore, 1986) obtained in previous experiments (Fig 3 from Dai et al., 1991).....  | 24 |
| <b>Figure 7.</b> Performance of one listener in three listening conditions, containing either 1 primary (700 or 1600 Hz) or both primaries (Fig. 2 from Macmillan and Schwartz, 1991).....   | 27 |
| <b>Figure 8.</b> A screenshot of the computer screen display response matrix for Experiment 1. The participant selected the interval which they thought contained a tone .....   | 41 |
| <b>Figure 9.</b> A screenshot of the computer screen display response matrix for Experiment 2. The participant selected the colour-number pair they thought they heard.....  | 46 |
| <b>Figure 10.</b> Percent correct scores, $p(c)$ , for Participant 3 in Experiment 1 (tone detection task). The open squares represent the conditions where only one frequency was presented. The closed circles represent the results from conditions in which both target and probe signals were presented. The data points for the T-only conditions are based on 168 trials per point. The data points for probe frequencies (900, 950, 1050 and 1100 Hz) in the T+P conditions are based on 24 trials per point. To simplify presentation of the results, the data point for the target in the two T+P conditions is the average of the performance for this signal in these conditions, and therefore this data point is based on 240 trials. .... | 49 |
| <b>Figure 11.</b> Percent correct scores, $p(c)$ , for Participant 7 in Experiment 1 (tone detection task). Symbols represent the same conditions as for Participant 3 (Fig 10) and are based on the same number of trials per point.....  | 50 |
| <b>Figure 12.</b> Percent correct scores, $p(c)$ , for Participant 12 in Experiment 1 (tone detection task). Symbols represent the same conditions as for participant 3 (Fig 10) and are based on the same number of trials per point.....   | 51 |
| <b>Figure 13.</b> Percent correct scores, $p(c)$ , for Participant 9 in Experiment 1 (tone detection task). Symbols represent the same conditions as for participant 3 (Fig 10) and are based on the same number of trials per point.....  | 52 |
| <b>Figure 14.</b> Mean Percent correct scores, $p(c)$ , for all participants in Experiment 1 (tone detection task). Each data point represents the mean across all participants for each frequency. The open squares represent the conditions where only one frequency was presented. The closed circles represent the results from conditions in which both target and probe signals were presented. To simplify presentation of the results, the data point for the target in the two T+P conditions is the average of the performance for this signal in these conditions. Error bars are $\pm 1$ standard error. ....  | 53 |

|  |    |
|--|----|
| <b>Figure 15.</b> Percent correct scores, $p(c)$ , for Participant 2 in Experiment 2 (sentence identification task). The open square represents the condition where sentences of only one $f_0$ were presented. The closed circles represent the results from conditions in which both target and probe sentences were presented. The data point for the T-only conditions is based on 168 trials per point. The data points for probe $f_0$ s (210, 215, 225 and 230 Hz) in the T+P conditions are based on 24 trials per point. To simplify presentation of the results, the data point for the target in the two T+P conditions is the average of the performance for this signal in these conditions, and therefore this data point is based on 240 trials. .... | 57 |
| <b>Figure 16.</b> Percent correct scores, $p(c)$ , for Participant 7 in Experiment 2. Symbols represent the same conditions as for participant 2 (Fig 15) and are based on the same number of trials per point. ....   | 58 |
| <b>Figure 17.</b> Percent correct scores, $p(c)$ , for Participant 4 in Experiment 2. Symbols represent the same conditions as for participant 2 (Fig 15) and are based on the same number of trials per point. ....   | 59 |
| <b>Figure 18.</b> Percent correct scores, $p(c)$ , for Participant 13 in Experiment 2. Symbols represent the same conditions as for participant 2 (Fig 15) and are based on the same number of trials per point. ....  | 60 |
| <b>Figure 19.</b> Mean Percent correct scores, $p(c)$ , for all participants in Experiment 2 (sentence identification task) as a function of $f_0$ . To simplify presentation of the results, the data point for the target in the two T+P conditions is the average of the performance for this signal in these conditions. The data point for the T-only condition at 220 Hz was displaced horizontally to more clearly demonstrate the results as compared to the data point at 220 Hz for the T+P condition. Error bars are $\pm 1$ standard error. ....   | 61 |
| <b>Figure 20.</b> Comparison of mean results of Experiment 1 and results of participant 0516 from the Greenberg and Larkin (1968) experiment. ....   | 68 |



## List of Tables

---

|  |    |
|--|----|
| Table 1. Characteristics of Conditions for Experiment 1: Pure Tones .....              | 43 |
| Table 2. Characteristics of Conditions for Experiment 2: Sentence Identification ..... | 47 |
| Table 3. Descriptive statistics for the five conditions for tone detection.....        | 56 |
| Table 4. Descriptive statistics for CRM sentence identification.....                   | 63 |

## Chapter One: Introduction

---

Every day we encounter an abundance of sounds. A mixture of sound sources can arrive at a person's ears at any time, and often only one source will be of interest. In many social settings, for example, a person must attend to a specific line of conversation while ignoring competing speech signals or other forms of background noise. The auditory system faces the problem of interpreting one complex waveform and having to separate this into spectral components to correspond with each different sound source. Once components are grouped appropriately, only those that are relevant in any given moment will be selected for further processing. This is called selective listening. It is an essential process for daily functioning, as it prevents the brain from being overloaded with irrelevant auditory information.

When multiple sounds are perceived simultaneously, they may have timbre, frequency, amplitude, and/or temporal differences. Psychophysical experiments have shown that listeners can use these differences in competing sound sources to isolate them as distinct entities. This form of acoustic analysis is called “bottom-up” processing. This phrase refers to the place where processing begins in the auditory system. Bottom-up processing begins low in the system, close to the periphery.

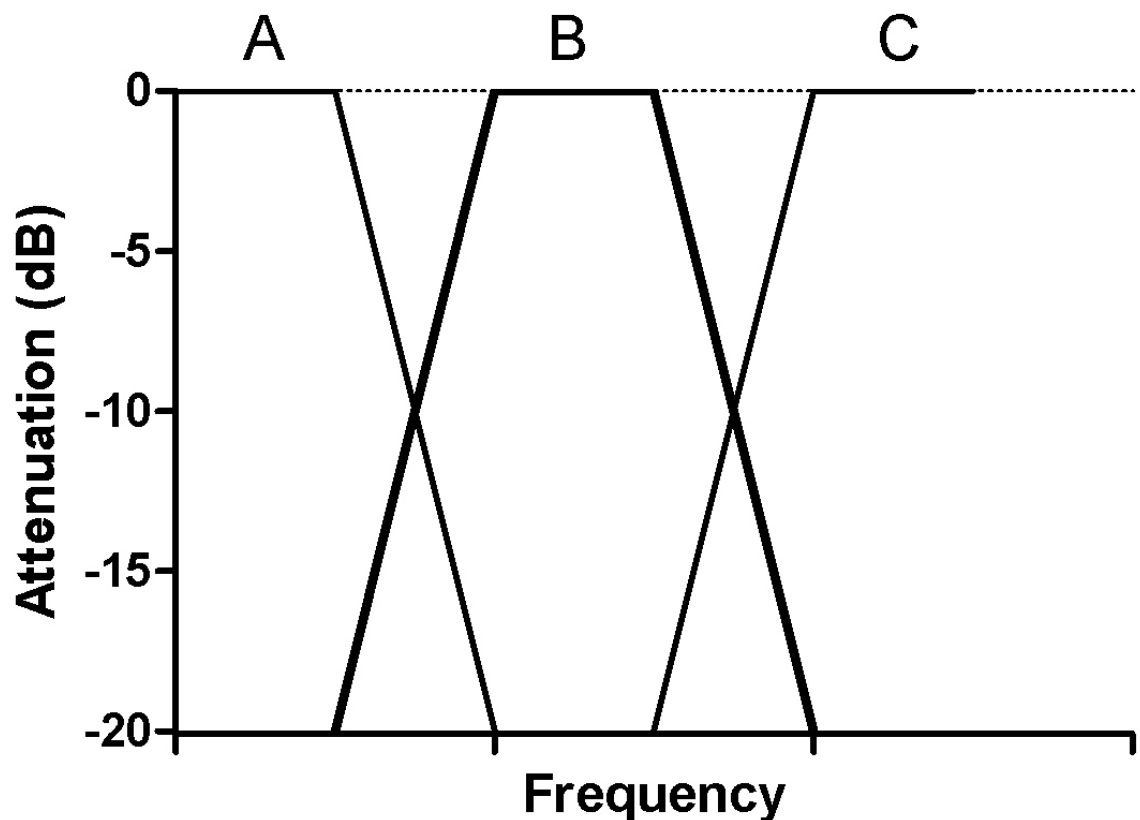
The auditory system can also use information from higher up in the system to process lower level information. These “top-down” processes include attention and learning, and can assist in selecting relevant signals from a mixture that includes irrelevant signals. This means that experience in listening to a particular sound can lead a person to expect a stimulus that matches the pattern that they are familiar with. For example, consider the scenario where a man attends a busy cocktail party with his wife. When his wife engages in a group conversation, his familiarity with the sound of her voice may allow him to better identify what she is talking about, as compared to identifying what a stranger in the group has to say. Indeed, experiments have been conducted which have identified that listeners can attend to

specific features of a repeated signal and use this stored knowledge to better detect that sound as it recurs.

Mechanisms of attention triggered by physical acoustic cues, such as those mentioned above, assist a person in developing an effective listening strategy for any given situation. The current study investigated low-level attention based on one such cue, fundamental frequency ( $f_0$ ) of voice, asking whether there is an "attentional filter" for  $f_0$ . The following section provides a review of the literature regarding the concept of the auditory filter, attention mechanisms in audition, auditory stream segregation, and the role of  $f_0$  for speech identification. A literature review of psychophysical techniques which can be used to investigate low-level attention mechanisms is also provided to support the formulation of the methodology for the current research.

### ***1.1 Frequency Selectivity, the Critical Band and the Auditory Filter***

The behaviour of the auditory system is often likened to that of an array of overlapping band-pass filters (Moore, 2012). Generally, a filter can be described as an electronic linear device which can be used to manipulate the spectrum of a signal (Moore, 2012). Filters can be low-pass, high-pass, band-pass or band-reject. Each type has different characteristics regarding how they attenuate certain frequencies. Whereas a low-pass filter will attenuate frequencies above a certain cut-off, a high-pass filter will attenuate frequencies below a cut-off. Band-pass and band-reject filters have two cut-off points and the frequency components between these will either be passed or attenuated. The typical characteristics of some of these filters are illustrated in Figure 1. Emphasis here is on band-pass filters. A bank of these are assumed to exist in the peripheral auditory system (Fletcher, 1940; Moore, 2012) and have a role in auditory frequency selectivity.



**Figure 1.** Characteristics of a low-pass filter (A); a band-pass filter (B); a high-pass filter (C)

Fletcher (1940) was the first to speculate, that when detecting a signal in background noise, listeners make use of filter pass-bands centred at the signal frequency on the basilar membrane (BM) of the cochlea. In this classical experiment, the bandwidth of a band-pass masker was increased, and the threshold for detecting a pure tone signal was measured (Fletcher, 1940). The noise remained centred at the frequency of the signal, and the spectrum level was fixed. For the narrowest bandwidths, the signal threshold increased as a function of masking noise bandwidth. At some point, however, the signal did not become less detectable with increased masking bandwidth, and the threshold remained constant. When the noise bandwidth is less than the auditory filter bandwidth, increases in noise bandwidth allow more noise to pass through the auditory filter. Once the noise bandwidth exceeds the auditory filter

bandwidth, however, no additional noise will pass through the filter with increasing noise bandwidth (Moore, 2012). Fletcher termed the bandwidth of noise which is critical to result in no further increase in signal threshold, as the “critical band”. In more recent studies, this term is often replaced by the term “auditory filter”.

Since the initial description of the auditory filter, there have been various methods derived to determine its shape. One clever method first described by Patterson (1976) involves using notched noise. Notched noise refers to noise from which a narrow band of frequencies has been removed. For this method, a fixed-frequency signal is presented in the presence of masking noise which contains a spectral notch centred at the signal frequency. The threshold for the signal is then determined as a function of varying notch-width. With increasing notch-width, up to the width of the auditory filter, there is less masking noise passing through the auditory filter, and so the threshold decreases. The presence and shape of auditory filters has been confirmed by more recent experiments involving notched noise (Zhou, 1995; Moore, Peters & Glasberg, 1990; Baker & Rosen, 2006).

The idea that auditory filters have a role in limiting auditory frequency analysis is now well-established, and a theoretical discussion has grown from this regarding the presence of specific filters which might select information which matches their pass-bands for further processing. Different terms have been used to describe this concept. These include: frequency response characteristics, listening bands, attention bands, and attentional filters. The term which has been selected for use here is ‘attentional filter’.

## ***1.2 Attention***

It might be considered that everybody has some idea of what “attention” is. Defining it, however, is by no means simple. Over the past century, researchers in numerous fields have contributed an overwhelmingly large body of research to the concept of attention. There

are several different types of attention (for review; see Moray, 1969); however, the focus here is on selective attention, and in particular, its relevance to auditory attention. The term “attention” in audition can in itself have different meanings. Attention can be either conscious (requiring higher-level processing), or unconscious (occurring at lower levels of processing).

It seems that humans are not capable of attending to all aspects of auditory input at one time. Thus, certain inputs are selected for analysis. The field of auditory attention is primarily concerned with the way in which incoming acoustic information is selected for further processing. Studies on the processing of sounds traditionally focused on bottom-up auditory analysis. More specifically, psychoacousticians looked at how pitch, intensity, duration and direction might influence how acoustic stimuli are perceived. Since the 1950s, however, there has been more of a focus on how top-down processing can affect how we perceive and interpret everyday sounds (Haft, Sarampali & Loui, 2007). There is a considerable amount of debate in the literature, and a multitude of theories, concerning where in the information processing stream attention begins to influence the analysis and segregation of auditory signals.

### **1.2.1 Theories of Auditory Attention**

Three influential theories on attention have been selected for discussion here. Each theory attempts to explain how attention can influence information processing. These include Broadbent’s Filter Theory (1958), Treisman’s Attenuation Theory (1960) and Deutsch and Deutsch’s (1963) late-selection account of attention. The former two theories are “early-selection” theories, suggesting that attention operates early in the line of information processing. The latter (as the name implies) is a late-selection theory which proposes that low levels of auditory processing are unaffected by attention, whereas later stages, such as the processing of semantic information, do require attention.

In Broadbent's comprehensive theoretical account of auditory attention, he described what is now cited throughout the literature as Broadbent's Filter Theory of Attention (1958). This theory proposes the presence of a selective attentional filter that allows a portion of total auditory input to go forward for further interpretation by more central processes (the attended message), whereas other portions are left in short-term memory to decay (unattended messages). In other words, an attentional filter can select the most relevant auditory information in the environment, based on acoustic properties such as frequency to go forward for further processing. These attentional filters may be activated based on the whether or not the features of an auditory signal match those which have previously been attended to. This increase in activity within relevant frequency channels (or auditory filters) can increase the signal-to-noise ratio (SNR) to aid signal detection by rejecting information from activation at more distant frequency regions.

The attentional filters described by Broadbent are not specific to frequency. Instead, they can be used to describe attentional selection in other auditory dimensions, such as signal duration (Dai & Wright, 1995) and frequency of amplitude modulation (Wright & Dai, 1998). They have been classified "all-or-none" filters (Haft et al., 2007) on the basis that some channels are determined as being irrelevant based on their underlying physical characteristics. Because of this, information in these channels does not go forward for further processing. Given that attention begins influencing how auditory information is processed at relatively low-levels, this is a classic example of an early-selection theory of auditory attention.

Broadbent's filter model suggested that unattended inputs are barred from entering conscious awareness. In other words, they are not processed for recognition or meaning. These propositions were followed by research which demonstrated that, for language, there are some semantic characteristics which are processed at a lower level. Experiments have been conducted requiring participants to attend to one of two distinct messages, each to a

different ear. They are then asked to repeat back one of these messages; this is called a shadowing task. It was discovered that participants switched their attention to the unattended message when that message contained their own name (Moray, 1960), or when it contained words that were relevant to the attended message (Treisman, 1960). Treisman (1964) revised the early-selection theory in order to explain these apparent interruptions to the all-or-none attentional filters proposed by Broadbent. Treisman suggested that instead of blocking unattended information, the filters selectively attenuate these inputs, making it difficult but not impossible for higher-level processing of the content that was attenuated.

Conversely, others propose a late-selection theory. This view posits that information in all channels is processed to the semantic level; for review, see Deutsch and Deutsch (1963). Accounts based on late-selection distinguish that form and meaning are extracted at the earliest stages of information processing (Duncan, 1980), as compared to early-selection accounts, which propose that only physical properties are processed at these early stages. For a full account of the theories of selective attention, see Driver (2001).

Currently, there is evidence in support of each theory mentioned above. However, Driver (2001) points out that with the vast body of data emerging from neuroscience, these basic theories which are so frequently provided in psychology literature are no longer adequate to explain mechanisms of attention. Much more complicated computational models are expected to be established in coming years. Despite this, some of the general principles of these early theories remain clear in biological findings. As an example, Driver points out that Broadbent's themes of "the selection of relevant stimuli", and "limited capacity" are consistent with current brain research (e.g. Alain & Woods, 1994; Da Costa, van der Zwaag, Miller, Clarke & Saenz, 2013). Additionally, Driver (2001) mentions that there is evidence emerging from neuroscience research which indicates that perceptual coding is modulated by



selective attention at quite early levels of processing, which is contrary to the ideas underlying the late-selection theory.

Studies of auditory attention tend to focus on models which measure a person's ability to detect or discriminate a signal of interest from background noise. In this context, attention can be understood to reflect a listener's ability to monitor some dimension along which a signal is expected to occur (Haft et al., 2007). This allocation of attention will often occur at early levels of processing, or in other words, before reaching conscious awareness. Thus, the meaning of attention which will be addressed in the current study does not refer to conscious attention – but to a pre-conscious, or lower level of attention.

### ***1.3 Auditory Stream Segregation***

There are two ways in which simultaneous auditory stimuli can be perceived. A subset of components may be grouped and perceived as though they come from a single source (called fusion). A different subset may be perceived as coming from multiple different sound sources (called fission; van Noorden, 1975). The latter process has also been termed auditory stream segregation (Bregman, 1990). Rogers and Bregman (1993) suggest that the auditory system begins by assuming fission (the presence of a single sound source) and once sufficient evidence builds to contradict this, fusion is established. This process, termed “streaming”, involves the coherent perception of a group of sounds and attributing it to one source. The way in which this is done will influence how these perceptual descriptions move forward for higher-order processing. The following section attempts to address some of the theoretical and experimental approaches to explaining what happens when a person is confronted with sounds from multiple auditory sources.

#### **1.3.1. Cherry's Description of the “Cocktail Party Effect”**

In 1953, Cherry described what is known as the “cocktail party effect”. That is - normal listeners have an ability to extract some features of the auditory stream, while ignoring other stimuli (such as background noise or competing speech). Cherry used a series of dichotic listening experiments to reveal how this perceptual parsing of different acoustic sources into different streams can be accomplished. In these experiments, subjects listened to a message over headphones, and reported what they heard, while another message by either the same talker, or a different talker, was played simultaneously. The subject’s performance was affected by two important factors. One factor was whether or not the messages were played to the same ear or to different ears simultaneously. When the target message was played to the opposite ear than the distractor message, performance was enhanced. Another factor was the predictability of the content of the message. When the participants could predict what the second half of the target message was from its relevance to the first half, they were better able to repeat the words. Cherry described the act of separating and attending to one message while ignoring another as “filtering”, the same idea later extended by Broadbent (1958).

As well as these two key findings, Cherry also mentioned other factors which could contribute to auditory stream segregation. He proposed that physical cues such as voice quality, and differences in the speed of speech could assist in the task. Also mentioned were perceptual cues such as voice pitch, and the perceived spatial location of the two sounds. Following this, Speith, Curtis and Webster (1954) conducted an experiment which aimed to determine which conditions could facilitate a listener’s ability to separate and attend to one voice message while ignoring an irrelevant message. To test this, the listener’s ability to answer one of two messages was measured. Manipulated variables included: horizontal separation of the sound sources in space, visual cues, and the aural shaping of the spectrum of the messages using filtering. Performance was enhanced for messages that were spatially

separated horizontally; however, visual cues did not appear to aid performance. Furthermore, it was demonstrated that when one message is high-pass filtered above 1600 Hz, while at the same time the other message is low-pass filtered below that frequency, segregation of the two messages becomes easier. Mean performance increased from 66% to 86% of answers correct with the addition of filtering. This result is evidence for the role of frequency separation in aiding auditory stream segregation.

Cherry (1953) also found that the participants could not identify the words or phrases of the unattended message. Furthermore, participants did not recognise when the unattended speech was reversed, or when it changed to another language. These results are consistent with Broadbent's filter theory (1958). Recall that his theory proposed that a selective attentional filtering mechanism would allow one of two simultaneous messages to move forward for further processing, while the other would decay in short-term memory. In contrast, a change of voice in one ear or another from male to female or vice versa, was nearly always identified by Cherry's participants, and thus he concluded that the ability to separate a speech signal from background noise is affected by factors such as the pitch of the speaker (Cherry, 1953).

Cherry's experiments (1953) utilised spatial separation as a cue for auditory stream segregation. Dichotic listening tasks involved presenting a sound or message to one ear, while at the same time presenting a competing sound or message to the opposite ear. The effects being studied were those of interaural time,  $f_0$  and level differences for determining the source of sounds. In a real-world listening environment, however, a person most often receives sounds from the same source to both ears and would not be able to make use of such interaural cues (Yost & Sheft, 1993). Therefore, these experiments in which sounds are delivered exclusively to one ear do not reflect the acoustics which occur in our everyday listening situations. What these experiments do highlight, however, is that auditory stream

segregation draws on both top-down and bottom-up auditory mechanisms. Cherry (1953) demonstrated that sound perception can be modulated by factors such as message content, and more primitive features such as source direction and frequency.

### **1.3.2. Bottom-Up versus Top-Down Segregation**

Bregman (1990) makes an important distinction between two forms of segregation. The first is “primitive segregation”, which is a bottom-up, pre-attentive process whereby sounds are segregated and grouped on the basis of auditory cues. Other literature has referred to this process as a sensory partitioning mechanism (Bey & McAdams, 2002). The second is “schema-based segregation”, which includes top-down attentive, or learning processes, arising from experiential and cognitive factors, involving voluntary or active listening. Other researchers have defined these two processes slightly differently. Cusack, Deeks, Aikman, and Carlyon, (2004) point out the difference between “selective attention” and “perceptual grouping”. They mention that in order for a listener to selectively attend to a particular sound; prior knowledge of the primitive acoustic properties of the target of interest is required. It is therefore argued that selective attention should be thought of as a top-down process, under conscious control of the listener. Furthermore, the authors highlight that performance is context-dependent, changing with a person’s degree of listening experience with the sound of interest. Perceptual grouping, on the other hand, should be thought of as an automatic, habitual process, taking advantage of the likelihood of acoustic properties of simultaneous sounds in order to group these into streams (Cusack et al., 2004). Perceptual grouping can occur irrespective of the listener’s context.

Bregman acknowledges that the contribution of primitive scene-analysis processes for the perception of speech can be obscured by schema-based processes. However he introduces auditory stream segregation as a preliminary process, and focuses his discussion on the role of primitive processes for the separation of speech sounds. One observation noted is that the

auditory system groups low-level properties and extracts these to develop a perceptual description of an individual's voice (Bregman, 1990). Among many other factors, Bregman pointed out that fundamental frequency (or the perceptual correlate, pitch) is one of the low-level stream-determining factors of sound which contributes to the segregation of sounds into individual percepts. This is the topic of the next section.

## **1.4 *Fundamental Frequency***

The perceived pitch of a person's voice is correlated with the physical feature of fundamental frequency ( $f_0$ ) of vocal fold vibration. Compared to women, men have larger and thicker vocal folds which vibrate at a slower rate, resulting in a lower  $f_0$ , and thus a lower perceived pitch. The average male speaker has an  $f_0$  of 125 Hz, whereas the average female speaker has an  $f_0$  of 225 Hz (Gussenhoven, 2004). A reduction in pitch occurs with a decline in subglottal pressure, and thus a release of vocal fold tension, whereas an increase in pitch occurs with the contraction of the cricothyroid muscles, and therefore the stiffening and lengthening of the vocal folds (Atkinson, 1978). The rising and falling pattern of pitch is referred to as the contour.

$F_0$  is not completely constant in on-going speech. Gaps need to be allowed for between words, and there are inherent differences in  $f_0$  between speech sounds (Cruttenden, 1986). Not all speech sounds require vocal fold vibration. Additionally,  $f_0$  can change within sounds and with stress in running speech. In order to use  $f_0$  as an acoustic cue for speech reception in noise, perceptual allowances are made for these gaps and changes in the  $f_0$  contour. In other words, the overall pitch pattern must be recognised.

### **1.4.1 The Role of $f_0$ on the Ability to Identify Speech in Noise**

The auditory system faces the common problem of receiving multiple speech signals overlying each other from different sources. In order to form a separate view of the different

spectra, the auditory system can segregate two steady-state vowels with different  $f_0$ s (Haukia & Bregman, 1984a; 1984b; Assmann & Summerfield, 1990; 1994). A psychophysical method called the double-vowel experiment has been used to investigate this effect. This involves having subjects identify both of a pair of simultaneously presented isolated vowels. Performance is quantified as the percentage of trials in which listeners can correctly identify both vowels in the double-vowel pairs. Consistently, identification is more accurate when the vowels are synthesized to have different  $f_0$ s than when they are the same (Bregman, 1990; Assmann & Summerfield, 1990; 1994; Arehart, King & McLean-Mudgett, 1997). These same mechanisms might also be used to derive multiple pitches and thus identities of voiced sounds in more natural speech.

Consistent with this, Broadbent and Ladefoged (1957) hypothesized that there is a pitch-detection mechanism which has a role in segregating and grouping overlapping formants. Listeners were presented with synthesized sentences to each ear containing the phrase “what did you say before that”. Sentences were generated and manipulated with a speech synthesizer which produced speech-like sounds (Broadbent & Ladefoged, 1957). The  $f_0$  could be changed independently for the first (F1) and second (F2) formants. This allowed sentences to be created in which the  $f_0$  on F1 was the same as or different from the  $f_0$  on F2. These were presented simultaneously either to separate ears or to the same ear. In one condition, F1 and F2 were presented to the same ear. In another condition, F1 was presented to one ear while F2 was presented to the other ear. Listeners were asked to determine whether there were one or two talkers present. The experimenters proposed that listeners would be able to combine the appropriate formants in speech based on similarities in  $f_0$ , in order to determine the presence of two talkers. The listeners usually reported hearing two voices when the two formants’  $f_0$ s differed by 10 Hz (the only separation tested). This failure to fuse the two talkers occurred in both same-ear and separate-ear presentations. When the formants

were synthesized to have the same  $f_0$ , however, listeners tended to fuse the speech sounds, reporting that they heard a single voice. The experimenters interpreted this to mean that when two overlapping speech sounds contain formants with the same fundamental, listeners are able to bind these together. This work provided early evidence that  $f_0$  is a powerful cue in auditory sound segregation.

Cutting (1976) extended these findings and investigated how  $f_0$  can affect a listener's judgment of the presence of one versus two sounds. In this experiment, participants listened to formants presented to different ears over headphones. On some trials, the same pair of formants (e.g., "da") was presented to both ears but the signals were synthesized with different  $f_0$ s. The participants were able to detect the presence of two separate sounds when their fundamentals were separated by as little as 2 Hz. One interpretation of these results is that the listeners were able to use the differences in  $f_0$  to segregate the sounds coming to the left and right ears and derive the identity of that speech sound.

Brokx and Nooteboom (1982) investigated how people can use differences in  $f_0$  to increase the intelligibility of sentences masked by other speech sounds. The listener's task was to repeat back Dutch nonsense sentences (target) which were presented simultaneously with another speaker reading a story (masker). The experimenters synthesized the target sentences in order to produce a constant  $f_0$ . Masking sentences were created to have either the same  $f_0$  as the targets, or an  $f_0$  of up to 10 semitones difference. A key finding was that as the difference in  $f_0$  between the speakers of the target and masker sentences increased, the intelligibility of the target sentence improved. When the target and masker  $f_0$ s were 3 semitones apart, the word recognition score was 65%. When the target and masker  $f_0$ s were equal, however, the word recognition rate was only 45% (Brokx & Nooteboom, 1982). In other words, the nonsense sentences were more difficult to identify when they were closer in

pitch to the interfering story. This result is consistent with what might be expected if the auditory system is able to segregate the auditory scene on the basis of pitch differences.

#### **1.4.1.1 The Processing of F0 and the Effect of Cochlear Hearing Loss**

Cochlear hearing loss can contribute to difficulties with pitch perception (Moore & Carlyon, 2005; Moore & Peters, 1992). More specifically, one of the physiological changes in cochlear functioning which is associated with hearing loss is the broadening of auditory filters. This results in reduced frequency selectivity, and associated difficulties with frequency discrimination. Studies have found that subjects with cochlear hearing impairment have a reduced performance on frequency difference limen tests (Simon & Yund, 1993; Moore & Peters, 1992). For example, Moore and Peters (1992) reported that hearing impaired listeners had increased frequency limens for all pure-tone frequencies ranging between 50 and 5000 Hz, compared to normal hearing listeners. Of more importance to the proposed research, f0 difference limens can also be affected by cochlear hearing loss. Moore and Peters (1992) found that subjects with hearing loss were significantly worse at discriminating f0s (ranging from 50 to 400 Hz) in complex tones than were subjects with normal hearing. Additionally, others have found a direct association between broader-than-normal auditory filters and larger-than-normal difference limens for f0 (Bernstein and Oxenham, 2006).

In a series of studies, Arehart and her colleagues investigated how adults with hearing loss compare against normal hearing adults in their ability to take advantage of f0 differences in double-vowel tasks (Arehart, 1998; Arehart et al, 1997; Arehart, Rossi-Katz, & Swensson-Prutsman, 2005). In the first task, double-vowel identification was measured as a function of differences in f0, denoted “delta f0”, or  $\Delta f_0$ , ranging from 0 to 8 semitones apart (Arehart et al., 1997). Results showed that both groups obtained a significant but similar amount of  $\Delta f_0$  benefit. For example, a  $\Delta f_0$  from 0 to 2 semitones improved overall performance by 18.5% for listeners with normal hearing, and 16.5% for listeners with cochlear hearing loss.



Group differences did emerge, however, from the results of the second task involving masked-vowel identification (Arehart et al., 1997). Performance in the masked-vowel task was measured by adaptively varying the SNRs. This was done in such a way as to estimate the lowest SNR required for 71% correct vowel identification in the presence of a masking vowel. There was a large mean improvement with  $\Delta f_0$ s for the masked-vowel identification task for both groups. Whereas the improvement for hearing impaired listeners was 4.4 dB, normal hearers showed a 9.4 dB improvement, almost twice that which was obtained for listeners with hearing loss (Arehart et al., 1997). These results suggest that listeners with hearing loss are less able to take advantage of  $f_0$  cues when listening for a particular sound which is masked by other sounds.

A similar pattern of results has been observed using sentence materials. Summers and Leek (1998) examined the effect of  $f_0$  differences for concurrent sentence identification. Participants included both hearing-impaired and normal hearing listeners. The target sentence had a fixed  $f_0$  of 120 Hz. There were five possible  $f_0$ s for the competing sentences, each presented in its own condition. In the first condition, the masking sentences had the same  $f_0$  as the target. In following conditions, they were manipulated to have an  $f_0$  either 2 or 4 semitones above or below that of the target. Both groups of listeners performed poorest in the condition where the target and masker had the same  $f_0$ . Normal hearing listeners showed a systematic increase in performance as the  $f_0$  difference increased from 0 to 2 and from 2 to 4 semitones. Hearing-impaired listeners had improved performance as the  $f_0$  difference was increased from 0 to 2 semitones, however, there was no further increase in performance with increasing  $f_0$  separation between the sentences (Summers & Leek, 1998).

It is now well-established that  $f_0$  is a valuable acoustic cue for benefiting speech identification in noise. Furthermore, when listening to competing sentences, normal hearing listeners achieve greater benefit from  $f_0$  differences than do listeners with a hearing loss (e.g.,

Summers and Leek, 1998). This disadvantage for hearing-impaired listeners could reflect a deficit in the basic processing of spectral/temporal information (bottom-up processing). However, a deficit in more top-down processes cannot be ruled out. The underlying reason may have implications for the development of speech-processing strategies in hearing aids and cochlear implants (CIs). Various methods have been proposed to enhance pitch perception in hearing aid and CI wearers, though all have shown limited functional benefit (McDermott and McKay, 1994; Green, Faulkner, Rosen & Macherey, 2005; Carroll, Tieden & Zeng, 2011).

## **1.5 *Detection of Tones in Noise***

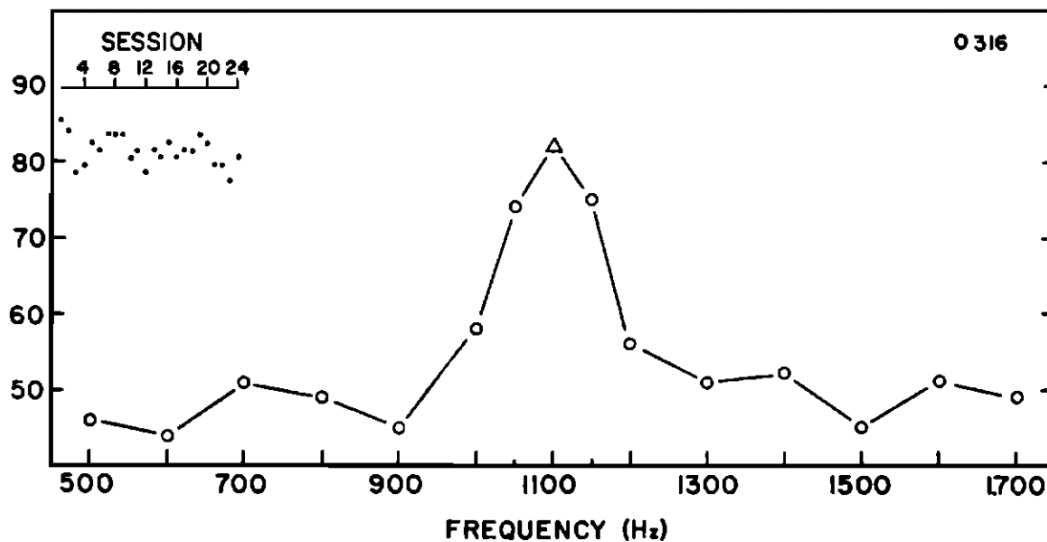
Listeners who have experience in detecting a signal of a certain frequency show poorer detection of signals which have the same energy, but differing frequencies. Perhaps the first experimenters to examine this phenomenon experimentally were Tanner and Norman (1954). They used a four-interval forced choice (4IFC) procedure to measure listeners' ability to detect a tone partially masked by noise. Their task was to select one out of the four noise intervals which they believed contained a tonal signal. A 1 kHz tone was presented as the target, set to a level at which listeners were able to correctly identify it 65% of the time. After several hundred trials, the experimenters changed the frequency of the tone to 1.3 kHz. As a result, performance declined to 25% (chance levels). However, when the experimenters informed the participants of the change in signal frequency and allowed them to listen to the new tone alone, performance quickly rose back to 65% (Tanner & Norman, 1954). This demonstrates that auditory attention can be tuned to a specific frequency region. Additionally, when listeners selectively attend to a narrow frequency band, their sensitivity to other frequency regions is decreased. This is consistent with the presence of an attentional filter tuned to the expected frequency.

### **1.5.1 The Use of a Probe-Signal Method to Estimate the Attentional Filter**

In order to measure the width of the attentional filter, Greenberg and Larkin (1968) developed a method of testing called ‘the Probe-Signal Method’. This involves first training subjects to detect signals of a single frequency in wideband noise. To maintain the subject’s expectation of this signal, it is presented on the majority of the sound trials. This is referred to as the “target” signal. “Probe” signals at different frequencies are then inserted randomly on the remaining trials. The probe-signal method utilises the two-alternative, temporal, forced-choice (2ATFC) procedure, whereby subjects indicate which of two randomly-chosen intervals contained the signal. In the classic experiment by Greenberg and Larkin (1968) the target and probe tone amplitudes were altered to result in equal detectability of each stimulus type when they were presented alone. This notion is demonstrated in Figure 2, where signals of each frequency are detected at the same rate whether presented as targets or probes. In other words, signals across frequencies were presented with “equally effective” amplitude (Greenberg & Larkin, 1968). Furthermore, an SNR was chosen at which the signal is detectable at a rate that is greater than chance but less than perfect. Using these techniques, Greenberg and Larkin (1968) found that the probability of detection was lower for probes than for target signals, and that the probability of detecting the signal decreased as a function of increasing distance from the target signal frequency.

Results are displayed by plotting the proportion of trials correct as a function of frequency, to reveal the “attention function” for each subject. The attention function demonstrates the basic effect of frequency selectivity. Greenberg and Larkin (1968) recognised a distinct similarity between the response patterns of the listeners and those characteristic of band-pass filters. An example of one participant’s attentional filter is demonstrated in Figure 2, where the expected (target) signal is 1100 Hz. The results show that the target signal was detected on the majority of trials, whereas the probe signal is

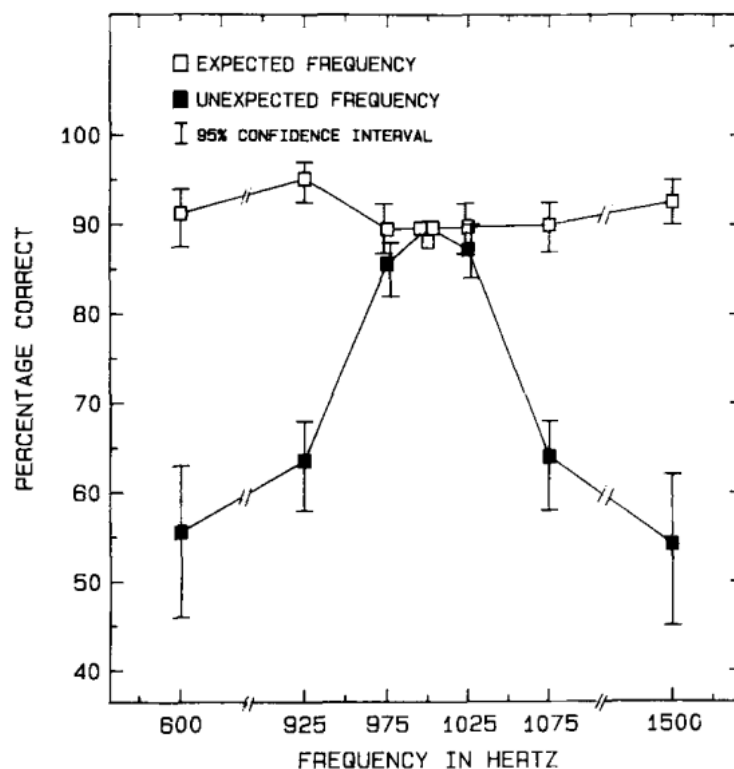
detected at near chance levels at about 150 Hz above or below the target signal. A similar pattern of results was found for other participants ( $n = 16$ ), and when using other target signals (e.g., 1000 Hz). The authors concluded that the results support a sensory filter model of frequency detection (Greenberg & Larkin, 1968). Another interpretation is that this pattern is a demonstration of the attentional filter for frequency.



**Figure 2.** The results of the probe-signal method for one subject's performance (percent correct) using 14 probe signal frequencies and a target signal of 1100 Hz. The closed dots represent the performance for probe signal frequency tones when they were presented alone as targets (Fig 2 from Greenberg and Larkin, 1968)

Scharf, Quigley, Peachey and Reeves (1987) conducted a series of experiments which extended the results of the classic Greenberg and Larkin study. They used the probe-signal method to measure the percentage of trials for which untrained subjects detected an expected, target frequency tone, as compared to probes with unexpected frequencies. They aimed to determine the bandwidth of the attentional filter. In other words, they looked at how different the probes needed to be from the target, in order for them to be detected near chance levels. Their finding was that when the frequency of the probe was greater than half a critical bandwidth away from the target signal, the detection rate dropped to approximately 65% (see Figure 3). When the probe frequency was more than two critical bands from the target,

detection rates dropped even further to chance levels (50%). Consistent with the results of Greenberg and Larkin (1968) the participants seemed to operate as though attending to a narrow band of frequencies centred at the target frequency. The authors suggest that subjects make choices about the presence of a signal on the basis of the anticipated frequency. The attentional filter can therefore ready the appropriate auditory filters prior to stimulation, to facilitate the detection of relevant stimuli (Scharf et al., 1987).



**Figure 3.** Mean percentage of trials on which the subjects correctly identified the interval that contained the signal. Detection hovers about 90% for expected signals across the frequencies; however detection rates fall away when the signals are unexpected probe

Recall that in order to measure the attentional filter shape, the probe-signal method involves presenting the target signal on the majority of trials to encourage listeners to attend to that frequency region. Scharf and colleagues (1987) conducted another experiment to determine the effect of repetition of the target signal on detection rates. To do this, they

presented a condition containing only probe signals (104 trials), and then repeated mixed conditions containing both probes and targets. In the probe-only conditions, the signals were detected at approximately the same rates as the target signals in mixed conditions. Moreover, even after being exposed to many probe signals (more so than the repetition of targets in mixed conditions), the probe detection rates remained poor in the mixed conditions (50%). Thus, the effect of repetition cannot explain the poor detection rates of probes in mixed conditions, nor can it explain the high performance for target detection (Scharf et al., 1987). The authors also emphasized that extensive training is not required for the study of selective auditory attention. It seems that listeners naturally and rapidly re-focus their attention to different frequency regions to enhance the detection of a new signal within the first few trials of a new condition.

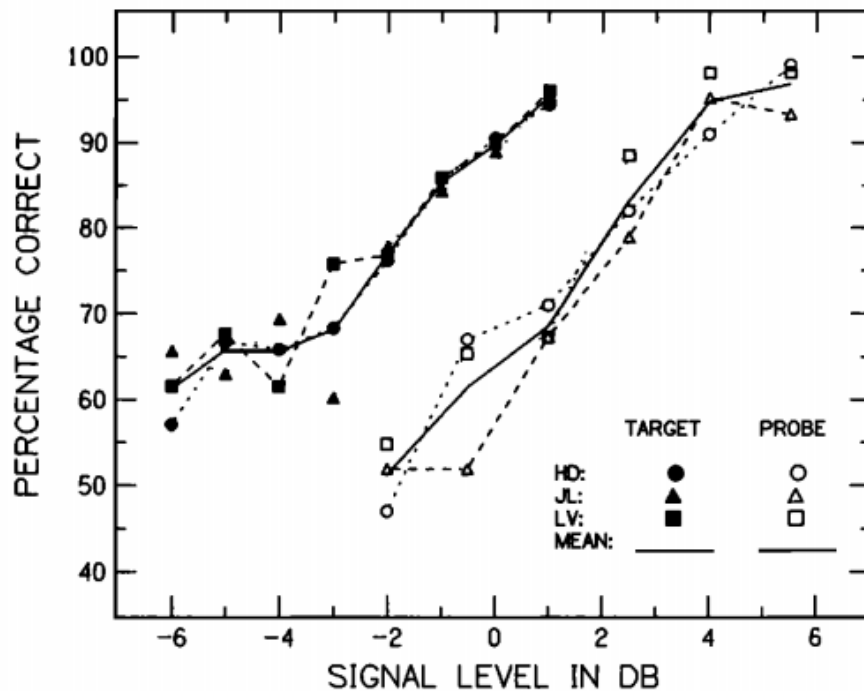
#### **1.5.1.1 Comparing the Attentional Filter to the Auditory Filter**

Different experimenters have different ways of illustrating the attentional filter. Whereas some researchers consider plots of performance in percentage correct against frequency as being indicative of attentional filter shape, others prefer to convert percentage correct into attenuation measured in dB as is done for measuring the auditory filter (Patterson & Moore, 1986; Dai, Scharf & Buus, 1991). This conversion allows researchers to compare the attentional band directly with the auditory filter bandwidth or critical band. A short review of the conversion process is discussed here (for a more detailed review; see Dai et al., 1991).

First, psychometric functions for both target and probe signals are derived for each listener. This involves measuring detection performance (in percentage correct) as a function of varying SNR for both targets and probes. This gives an estimate of the attenuation of probe signals. In other words, by what amount does the probe signal level need to be increased, for it to be detected at the same rate as target signals? An example of this process can be found in

a study by Dai et al (1991). Psychometric functions were measured for both a 1 kHz target frequency, and for probe signals of 24 frequencies separated at least 0.23 kHz from the target (10 between 0.25 – 0.77 kHz, and 14 between 1.27 – 3.4 kHz). Figure 4 shows the psychometric functions for three listeners with a 1 kHz target and the 24 probes (Dai et al., 1991). The functions obtained for the probes and target signals are roughly parallel, however there is an average attenuation of the probes of about 4 dB. In other words, the psychometric functions for the probes are shifted to the right, relative to the functions obtained for the target. This means that listeners required the probes to be presented at a greater level in order for them to be detected equally to target signals. The maximum attenuation of probes located outside the critical band centred at 1 kHz was estimated to be 7 dB. In other words, regardless of their distance from the target frequency, probe signals showed a maximum attenuation of 7 dB.

The attenuation provided by this attentional filter could then be compared to that which would be predicted by measures of the auditory filter. Dai and others (1991) found that the amount of attenuation produced by the attentional filters was significantly less than that which is produced by the auditory filter centred on the target frequency, which would be at least 20 dB for the probes used in their experiment (Dai et al., 1991). The authors suggested that listeners may make some use of the auditory filters in the regions of the probe signals, but perhaps assign these bands reduced weight as compared to those in target frequency regions (Dai et al., 1991). Regardless of the reason, the important contribution of these authors is that beyond the attended filter (at least at 1 kHz), the attention function is relatively flat (Dai et al., 1991).

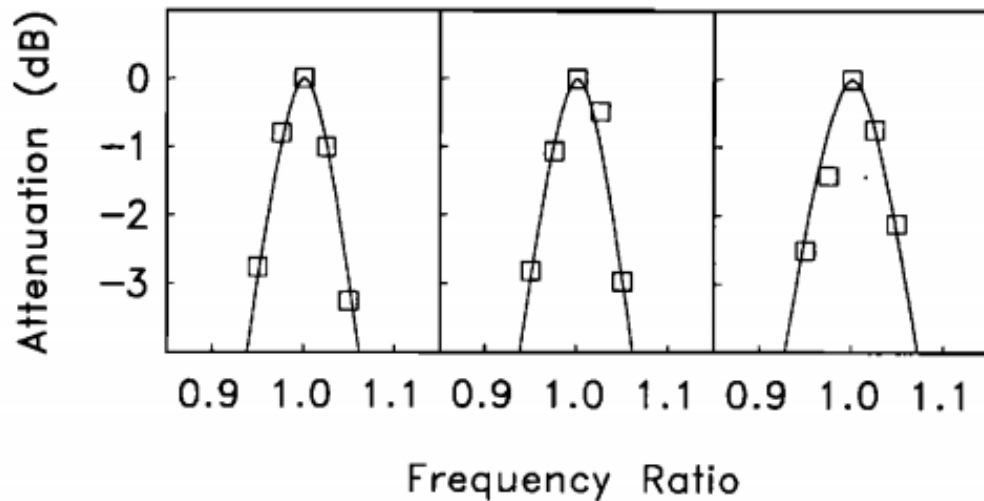


**Figure 4.** Listening bands/psychometric functions for a 1 kHz target (closed symbols) and probe tones (open symbols) derived using a probe-signal method for three participants (Fig. 4 from Dai et al., 1991)

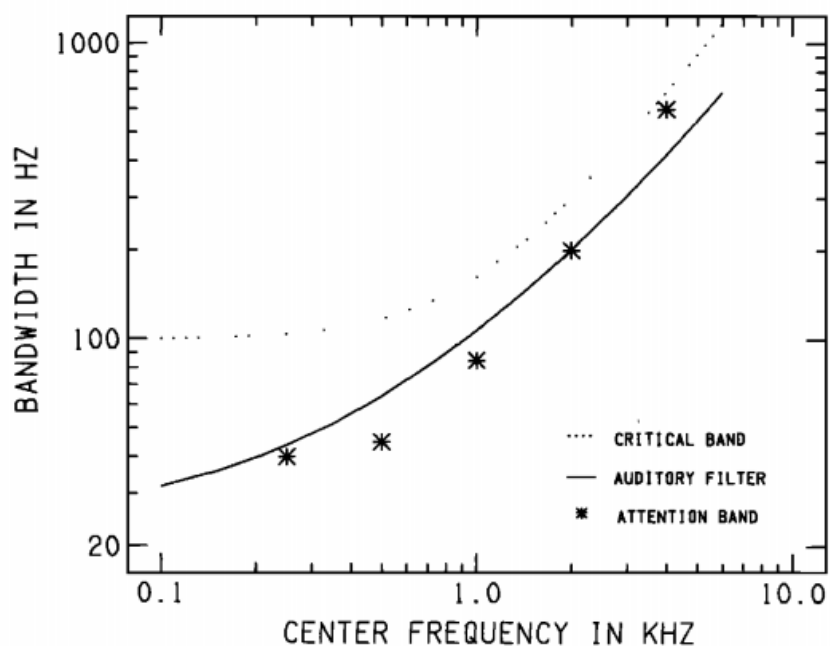
Once psychometric functions have been obtained, they can be used to derive listening bands plotted in attenuation (dB) as a function of frequency. An example of an attentional filter constructed in this way is shown below in Figure 5. Once this conversion is complete, researchers can then compare the attentional filter directly with the auditory filter bandwidth or critical band. In the same series of probe-signal experiments, Dai et al (1991) did this. They compared the mean derived attentional filters with auditory filters measured by Patterson and Moore (1986) and critical bandwidths measured by Zwicker and Terhardt (1982) for five frequencies (0.25, 0.5, 1, 2 and 4 kHz). They found that the attentional filter was similar to the auditory filter in both width and shape (see Figure 6), with only two small discrepancies. Firstly, at 0.5 kHz, the skirt of the attention function was steeper than that of the auditory filter function at the higher frequency end; and secondly, the attention function was broader than the auditory filter function at 4 kHz. Overall, however, there were marked



similarities between the three measures. This could indicate that attention is focused on the auditory filter (centred at the target frequency) during these tasks.



**Figure 5.** Example of attention functions plotted with attenuation (dB) as a function of frequency (Fig 4 from Schlauch & Hafter, 1991)



**Figure 6.** The bandwidth of the attentional filter (labelled attention band) derived from attenuation and plotted as a function of target frequency. Plotted for comparison are measures of the critical bandwidth (Zwicker & Terhardt, 1982) and auditory filter (Patterson & Moore, 1986) obtained in previous experiments (Fig 3 from Dai et al., 1991)

### **1.5.1.2. Single-Band Model versus a Multiple-Band Model**

The fact that auditory signal detection performance declines when listeners are uninformed about the frequency of an incoming signal, termed frequency uncertainty, has led investigators to question how attention is distributed across the frequency range when performing these detection tasks. Two categories of models have been postulated to describe the outcome of uncertain-frequency detection. The first model, first introduced by Tanner, Swets & Green (as cited in Swets, 1963) proposes the use of a single attentional filter which can be adjusted in location in accordance with the predicted frequency of a stimulus. In contrast, one of the first multiple-band models (introduced by Green, 1958) assumes that there are multiple attentional filters in use at any time, between which the input can be compared to make a decision about the presence of a signal. According to this model, the listener who attends to filters at various frequencies will be listening to more noise power from the masking noise which is located between those frequencies. However, they would not be listening to more signal power than that predicted by a single-band model (Swets, 1963). Moreover, the multiple-band model predicts that detection performance will decrease as the frequencies attended to increase in separation – until the point where the two bands no longer overlap. Overall, as compared to the single-band model, this predicts greater detection of signals with extreme changes in frequency from that which the listener is expecting (Swets, 1963).

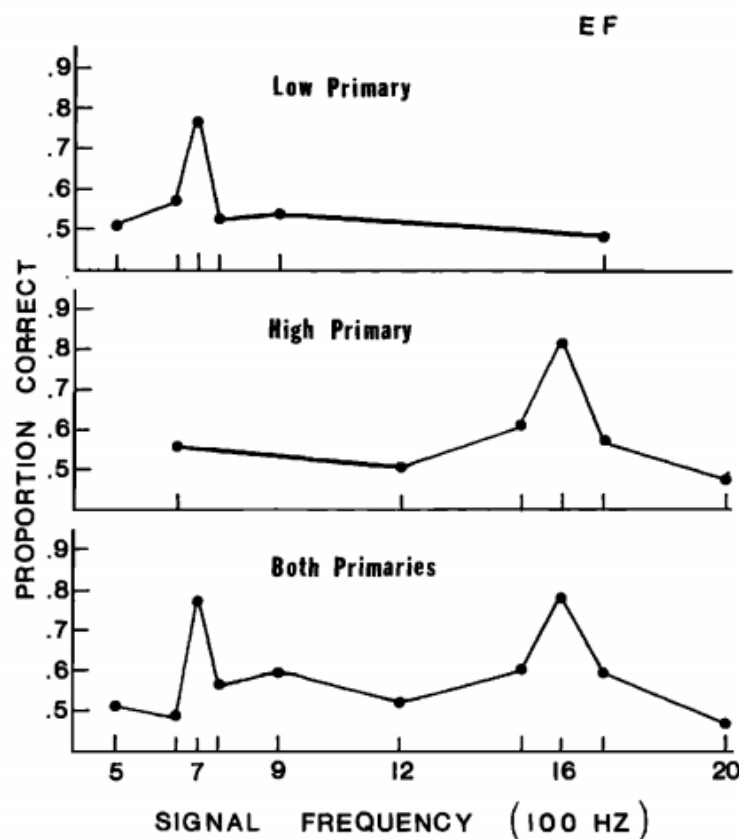
Perhaps the first experiment to attempt to fit data to a specific model was conducted by Tanner (1956). It was predicted that when two signal frequencies are sufficiently separated, it would become more difficult to attend to both frequency regions, resulting in decreased detection rates for those tones. To test this, Tanner (1956) conducted a multiple-session experiment with two participants which varied the frequency separation of two signals in a 4IFC procedure. The task required that the participant select one of four noise

intervals which they believed contained the tone. One of two different frequency signals of equal likelihood was presented on each trial. Different sets of frequencies were used in different conditions, ranging between 700 to 1300 Hz. Frequency separations ranged from 25 Hz to 600 Hz. The overall finding was that when the frequencies were separated by 100 Hz, detection rates began to decrease. At approximately 300 Hz separation, detection reached a minimum. This data is described as being in line with a single-band model, where a single monitored attentional filter is capable of being rapidly swept across the frequency range (Tanner, 1956). Thus there must be some limit as to how far the filter can sweep to facilitate detection of distant frequencies. Alternatively, the results could lend support to the multiple-band model. As the frequency separation increased, so might the power from the masking noise, which could result in the poorer detection rates.

Subsequently, Green (1961) designed an experiment which measured a listener's ability to detect auditory tonal signals of known frequencies compared to those which are selected randomly from a range of frequencies. This experiment showed that listeners are better able to detect a tonal signal in noise when that tone is fixed in frequency compared to when it varies on a trial-by-trial basis (Green, 1961). Hübner and Hafter (1995) termed this the "uncertainty effect". Another finding was that the loss of sensitivity to uncertain frequency signals was smaller than that which what would be predicted by an ideal multiple-band model (Green, 1961).

In a later study, Macmillan and Schwartz (1975) conducted an experiment utilising the probe-signal method to help clarify which model is most accurate. The experiment consisted of three conditions: a low-frequency (700 Hz) target condition, a high-frequency (1600 Hz) target condition, and a condition where either the low- or high-frequency targets could occur. Probe signals ranging between 500 and 2000 Hz occurred on 23% of the trials. The attentional filters for each of three listeners showed peaks of sensitivity at the target

frequencies; regardless of whether one or two targets were presented in any given condition (an example of one listener's performance is displayed in figure 7). In the two-target condition, sensitivity was reduced to near chance levels between the target frequencies, indicating that listeners can attend to two different frequency regions simultaneously without enhancing their sensitivity to signals of intermediate frequencies. If listeners can attend to more than one frequency band at a time, this is evidence against an adaptive single-channel attentional filter model (Macmillan & Schwartz, 1975).



**Figure 7.** Performance of one listener in three listening conditions, containing either 1 primary (700 or 1600 Hz) or both primaries (Fig. 2 from Macmillan and Schwartz, 1991)

It is clear from the literature reviewed here that no consensus has been reached regarding which model best describes the large body of data obtained by researchers so far. In some cases, the detection of unexpected frequency probe tones is better than that which would be expected if subjects monitored the output of a single auditory filter centred at the

target frequency. In other studies, sensitivity was greater than that which would be predicted from a multiple-band model. It could be that listeners choose either a single- or multi-band strategy depending on the task; however whether this is by automatic processing, or their own control, remains unknown.

#### **1.5.1.3. The Effect of Listening Strategies**

Penner (1972) found that participants changed their decision-making criteria and pattern of responses as the perceived rewards were manipulated in a probe-signal method experiment. The experimenters varied monetary pay-offs for tone-detection between the conditions to look for changes in the participants' listening strategy. The first condition was intended to motivate participants to attend to tones that had the same frequency as a cue tone. The second condition essentially instructed participants to attend to all frequency tones, even if they differed from the cue tone. Data from the first condition were similar to those of Greenberg & Larkin (1968), with the results of each participant displaying an attentional filter approximating the auditory filter shape, centred about the cue tone frequency. Results from the second condition revealed much broader shaped frequency responses, with poorer detection of the cue tone frequency, though better detection of distant frequency tones. The results demonstrate that participants can use different subjective strategies for pure-tone signal detection.

#### **1.5.1.4. Signal Uncertainty: Frequency Cueing and Listening Bandwidth**

Uncertainty can be varied by changing the number of possible frequencies from which a probe signal is drawn. A succinct explanation of this process is given by Hafter, Schlauch and Tan (1993). They describe frequency uncertainty by considering a case in which the to-be-detected signal has a frequency which is one of  $M$  possibilities. When listeners are familiar with the signal frequency prior to measurement, and when  $M = 1$ , performance is

best. As  $M$  increases ( $M > 1$ ), performance worsens. This is evident in studies showing that in order to produce detection rates of highly uncertain probes (drawn from a wide range of frequencies, e.g. 600-3750 Hz in Schlauch & Hafter, 1991) which is similar to that which is achieved when  $M = 1$ , that signal must be raised by 3 – 5dB (Green, 1961; Schlauch & Hafter, 1991). These studies illustrate the cost of sharing attention across a wide range of frequency bands.

The uncertainty effect can be compensated for by presenting clearly audible cues before the target signal (Mondor & Bregman, 1994; Tan, Robertson & Hammond, 2008). Tan et al (2008) employed a variant of the probe-signal method to investigate the role of frequency cueing with single pure tones. Three experiments were conducted. The first experiment was similar to that of Greenberg and Larkin (1968), though they presented a cue before each trial that was fixed to be the same as the target frequency. Results of this experiment were consistent with previous research, finding mean detection performance was best for the target frequency, and then declined with probe signal deviation from the target frequency. In the following two experiments, auditory cue frequencies were varied, and the effects were measured on the detection of probe frequencies. When they presented an auditory cue before the to-be-detected signal, they found that detection rates were better when the auditory cue matched that of the to-be-detected signal. Detection of the signal declined as the probe frequency deviated from the frequency of the cue. In summary, Tan et al concluded that two auditory mechanisms combine to produce these findings. Firstly, irrelevant frequency stimuli are suppressed based on developed expectations for the incoming signal. Secondly, the detection of the signal (whether it is a target or a probe) is enhanced with the presentation of a relevant auditory cue beforehand (Tan et al., 2008).

Schlauch and Hafter (1991) used different types of stimulus cues to see which would be effective for improving tone detection performance. Variations in uncertainty were

achieved by providing cues before each trial which consisted of one, or two or four tones presented simultaneously. On target trials, the cue tones had one component whose frequency matched the target signal. The authors called these “iconic cues”. These were presented on 75% of the trials. On remaining trials, probe signals were presented. The probe signal frequency differed from one component of the cue by a fixed ratio. Performance was best for tones following iconic cues, and declined as the ratio between the cue tones and the to-be-detected signals increased. Schlauch and Hafter (1991) proposed that iconic cues can stimulate the filter at the frequency location of the target tone which follows, thereby informing the listener of where to listen. The further the probe tones fall from the cue tones, the less likely it is that the filters in the probe region will be stimulated prior to the presentation of the probe.

In order to further investigate the role of various cues on signal detection, the same authors conducted a second experiment (Schlauch & Hafter, 1991). Psychometric functions (in percentage correct) from the first experiment were converted to effective signal levels (in dB) to estimate each participant’s listening bandwidth (Schlauch & Hafter, 1991). Recall that this is another term for the attentional filter. When the cue was only a single tone (which matched the target), the listening bandwidth was 12% of the centre frequency. This is roughly the same as the auditory filter widths which have been measured using the notched-noise masking method (Moore & Glasberg, 1983; Dai et al., 1991). When the cue contained four components (with one matching the target), however, the listening bandwidth was approximately 14% of the centre frequency. Although the difference was small, the authors interpreted this pattern of results to mean that at least some of the loss of sensitivity in tone detection which occurs with listening uncertainty can be accounted for by the widening of listening bands (Schlauch & Hafter, 1991).

#### **1.5.1.5. Are Probe Signals Heard but not Heeded?**

There has been some disagreement about whether or not the probe signals are in fact heard, but consciously rejected, or whether the listeners are simply not aware of their occurrence. The former conjecture is referred to as the “heard-but-not-heeded” hypothesis (Scharf et al., 1987). In a discussion of their experimental findings, Scharf et al (1987) suggested that some component of listener’s response patterns may arise from a heard-but-not-heeded strategy. This conclusion stems from results of probe-signal experiments 8 and 9 in the research by Scharf et al (1987). In these experiments, complex sounds were used as probe signals. In experiment 8, the target was a 1 kHz tone. The probes were 2-tone complexes separated by 90 Hz, and centred at either 1 kHz (probe A) or 1.2 kHz (probe B). The results were that the complex probe sounds were detected better when they fell within the critical band of the expected signal (mean detection was approximately 70% for probe A), than when they were outside this range (mean detection of probe B was 55%). In experiment 9, the target was a 400 Hz tone, and probes were complex tones with energy at three frequencies (1.6, 2 and 2.4 kHz). The frequency information from this tone complex can be processed to calculate the missing fundamental (Schouten, 1940) and so this is perceived as having the same pitch as a 400 Hz tone. Interestingly, the subjects were unable to detect the probe in experiment 9 (approximately 50% detection) despite it having a perceived pitch which was equal to that of the target tone. The authors suggested that this may have been due to greater qualitative differences between the sounds, which may have led the probes to be rejected as being part of the background noise. This is consistent with the heard-but-not-heeded hypothesis as it suggests that participants could hear the probes, however rejected them to prevent false alarms. Whether or not this rejection was made consciously or not remains unknown.



In another attempt to test the hypothesis that participants hear but reject unexpected stimuli, Scharf et al (1987) ran two additional experiments. Firstly, they informed participants that signals other than the target would be presented random trials. This did not improve their detection of distant probes. Secondly, the experimenters participated as subjects in the experiment themselves, knowing in advance exactly which probe stimuli were to be presented. It was found that the experimenters performed just as poorly on probe trials as did naïve participants. Scharf et al (1987) interpreted these results as being evidence against a post-stimulus response choice (conscious rejection of probes). Thus, the results could not be explained by the heard-but-not-heeded model.

This finding was replicated by Dai et al (1991). They found that when an experimenter, with knowledge about the stimuli, participated as a subject; this had little effect on the response patterns. Together, these results can be used to infer that there is no need to be ambiguous about whether the subjects should expect to detect signals of frequencies other than the primary signal. Participants can therefore be told to expect signals of varying frequencies before commencing any listening task involving the probe-signal method.

#### **1.5.1.6. Individual differences**

Other studies have investigated whether or not there is a relation between the response patterns of listeners and their individual characteristics. One which can play a role in frequency selectivity experiments is the hearing ability of the listeners. Moore, Hafter and Glasberg (1996) measured the shape of the auditory filter using a notched-noise masking technique described by Glasberg and Moore (1990) in four normal hearing participants and two participants with cochlear hearing loss. In a second experiment, the same participants performed tasks using the probe-signal method to measure attentional filter bandwidths. Results were analysed to look for associations with auditory filter shape. This was the first

experiment to directly compare auditory and attentional filters, each measured on an individual basis and compared *within* subjects.

The first experiment measured auditory filters for each subject at four centre frequencies ranging between 1 – 2 kHz. Overall results indicated that auditory filters for the impaired ears were broader than those for normal ears. In the second experiment, probe frequencies for each target were chosen to correspond to the points on each individual's auditory filters where the target tone would be attenuated by 5 and 10 dB (for normal ears) or 2 and 3 dB (for impaired ears). A lesser amount was chosen for impaired ears given that the greater amount corresponded to probes which would be extremely distant from the target, due to their broader auditory filter shapes (Moore et al., 1996). In order to compare the attentional filter directly with the auditory filter, another experimental condition was used in which only the target was presented (fixed in frequency) however it was attenuated by 0, 2, 3, 5, or 10 dB relative to the level used in the probe-signal experiment.

Results for one of the participants with cochlear hearing loss showed attentional filter results which were similar to that predicted by their auditory filter. However, the other participant in this group showed great asymmetries in their responding, with greater detection of probe signals at frequencies above the target, than of probes at frequencies below the target. A possible interpretation of these results, offered by Moore et al (1996) is that this participant had an asymmetry in their underlying auditory filter, which could not be measured in the first experiment because they were extremely broad. It may be that broad auditory filters mean that adding a signal does not produce a distinct peak in the excitation pattern, as it does for normal hearing listeners (Moore et al., 1996). Hearing loss might therefore lead to the monitoring of multiple auditory filters simultaneously, or in other words, a more “broadband” listening strategy (Moore et al., 1996). This might explain why listeners with

hearing impairment are more reliant than normal hearers on increasing the level of a signal in the presence of background noise to assist detection.

In young, normal hearing listeners, it seems that the detection of tonal signals in noise is more effective when features (such as the frequency) of the incoming signal are known and can be attended to. Ison, Virag, Allen and Hammond (2002) used the probe-signal method to measure attentional filters for both young and elderly listeners (with hearing thresholds typical for their age) to look for any effects of age. Ison et al (2002) predicted that the attentional filter width would increase with age, as is the case for auditory filters.

Interestingly, the attention functions of both groups were almost identical; each showing the similar rates of target detection and systematic declines in signal detection for probes as they deviated from the target frequency. Ison et al (2002) conclude that the frequency selectivity provided by the attentional filter is preserved in elderly listeners. The authors emphasize, however, that decrements might become apparent in more realistic listening conditions and suggest that more direct measures of speech recognition are required to understand its relation to selective attention (Ison et al., 2002).

The olivocochlear bundle (OCB) has been well described anatomically, but its function in hearing is not well understood. Scharf, Magnan and Chays (1997) investigated its role in 15 patients, before and after undergoing a vestibular neurectomy, using a series of 14 psychoacoustic and audiometric tests. Tests were grouped into four categories: detection (including selective attention), discrimination (of intensity and frequency), loudness and “other” (including lateralisation). A vestibular neurectomy is a surgery performed to alleviate severe attacks of vertigo, most often in patients with Ménière’s disease. The operation involves the sectioning of the vestibular portion of the eighth nerve, along with the OCB (Scharf et al., 1997). Comparison of pre- and post-operative measures revealed a clear change in hearing on just one set of results from within the extensive test battery. This was in the

detection of signals at unexpected frequencies as determined by a probe-signal method. Pre- and post-operative results for this task were available for four patients. For another eight participants, results obtained using the healthy ear were compared to those obtained using the operated ear. Generally, results revealed that patients perform better for the detection of tones with unexpected frequencies in the ear that no longer received efferent input. Prior to the surgery, these patients responded with the characteristic auditory filter-like response. The authors thought that the improvement in the detection of unexpected stimuli might indicate a loss of the ability to selectively respond to one frequency region (Scharf et al., 1997). This finding implies that the attentional filter for frequency operates at the periphery of the auditory system to exclude sounds with unexpected frequencies from central auditory processing. Furthermore, this frequency selectivity is, at least partially, established and maintained by the integrity of the OCB.

#### **1.5.1.7 Measures of Selectivity in Stimulus Domains other than Frequency**

There is some evidence that attentional filters may exist in dimensions other than sine wave frequency. One dimension to consider is the duration of the expected signal. Dai and Wright (1995) utilised the probe-signal method in two experiments to investigate this effect. In the first experiment, subjects were encouraged to expect a target signal of either 4 ms or 299 ms duration. The authors called this experiment: “Detection of signals having unexpected durations”. Four different target signals were used in separate conditions (0.25, 1 and 4 kHz, and noise). Probe signals with durations of 4, 7, 24, 86, 161 and 299 ms were then inserted on random trials. Signal detection dropped from approximately 80 – 90% to 50% as a function of deviating duration from that of the target (in a similar manner that is observed in the frequency domain) for all signal types. For example, when the target was a 4 ms signal, probes of 299 ms were detected at chance levels, however with decreasing probe duration,

detection rates steadily increased. The same pattern of results was found when the target was a 299 ms signal. In this condition, the 4 ms probes were the most poorly detected signals, and detection increased with increasing probe duration towards that of the target signal.

In another experiment, subjects were instructed to expect signals of the same range of durations which were used in the first experiment (Dai & Wright, 1995). This was called “Detection of signals having *uncertain* durations”. Subjects were given a preview of the duration of each possible signal (at a level 5 dB above those in the experimental trials) at the beginning of each block. Each signal duration was presented 10 times within a 60-trial block. Detection of signals in this experiment was on average 6% lower than the detection of the same signals in probe-only conditions. Thus, uncertainty about signal duration has little effect on performance. A possible interpretation of these results, offered by Dai et al (1995), is that listeners are capable of monitoring a range of signal durations at a time, and that they can adjust their listening strategy with respect to the demands of the auditory task.

In more recent research, Reeves (2013) altered the duration of multiple probe signals to investigate how this might affect the shape of the attention functions. In these experiments, a multi-probe 2IFC procedure was used with probes of three durations: 20 ms, 30 ms, and 300 ms. They found an unexpected asymmetry in the attention function for only the short duration probe tones (20 ms and 40 ms) with frequencies that were lower but close to the target frequency, as compared to the same duration tones that were higher but close to the target frequency. When the target was a 1000 Hz signal, there was approximately a 10% difference in the overall detection rates of a 925 Hz probe tone compared with a 1075 Hz probe. This asymmetry was replicated across five experiments, with subjects performing consistently better for low-frequency probes over high-frequency probes, and in some cases – over the target signal itself. This result seems puzzling, as why would a low-frequency probe signal be detected better than a signal for which attention is directed? The authors proposed

that there may be asymmetrical underlying filters which can explain the findings. This may mean that there is more attention directed to filters at some distance below the target frequency, than there is for those at the same distance above the target frequency (Reeves, 2013).

Selectivity based on amplitude-modulation of a signal has also been investigated using a probe-signal method (Wright & Dai, 1998). In this experiment, the target modulation rate was either 4, 32 or 256 Hz, and probe modulation rates included: 4, 8, 16, 32, 64, 128 and 256 Hz. When the target rate was at 4 Hz, there was about a 10% drop in the detection of modulation depth of all probe rate signals, as compared to target-alone conditions. Performance for 4 Hz modulation signals in conditions where they were presented alone was 91%. When the target rate was 32 Hz, and 4 Hz probe rates were presented, mean performance was down at 58%. Similarly, when the target rate was 256 Hz and the 4 Hz probe was presented, these were detected on average, 62% of the time. Despite these differences, however, when the target rate was either 32 or 256 Hz, and probe rates 16 Hz or greater were presented, there was no significant difference in detection rates compared to probe-alone conditions (Wright & Dai, 1998). Expecting a given modulation rate had some influence on the detection of modulation depth at unexpected probe rates; however the results were not consistent. In particular, unexpected modulation rates of 16 Hz or greater were not as poorly detected as one would expect if there were a sharply tuned attentional filter for this domain. The authors suggested that if filters tuned for modulation frequency do exist, then listeners may monitor multiple filters concurrently (Wright & Dai, 1998).

## ***1.6 Hypotheses and Rationale***

The idea that listeners have a selective attentional filter for frequency whose shape approximates that of the auditory filter has been relatively well-established (Greenberg & Larkin, 1968; Scharf et al., 1987; Dai et al., 1991; Dai & Wright, 1995). Research has also

tested for the presence of an attentional filter in other domains. As well as frequency, the probe-signal method has been applied to other acoustic cues. These include pure tones in noise that vary in signal duration (Dai & Wright, 1995) and frequency of amplitude modulation (Wright & Dai, 1998).

The current study thus aimed to replicate and extend previous results, by applying the probe-signal method to an additional stimulus dimension:  $f_0$  for speech stimuli. The study asked whether there is an attentional filter in the  $f_0$  domain, similar to that which has been found in the frequency domain. To test this, participants listened to sentences presented in background noise. Target sentences with a fixed  $f_0$  were presented most of the time. Probe sentences having differing  $f_0$ s were presented randomly on the remaining portion of sentences to see if this had an effect on sentence identification. If an attentional filter for  $f_0$  did exist, the consequence would be better identification of target sentences as compared to probe  $f_0$  sentences. Furthermore, identification performance should decrease as a function of increasing difference between the probe  $f_0$  and target  $f_0$ .

## Chapter Two: Methods

---

The current study investigated the effect of auditory attention and frequency or  $f_0$  on the detection of tones and the identification of speech in the presence of background noise. The study consisted of two experiments completed in a single session lasting up to two hours. Both experiments employed a two-alternative, temporal forced-choice (2AFC) probe-signal method, following the procedure of Greenberg and Larkin (1968). A critical feature of this method was for each listener to first develop an expectation for the target stimulus (Greenberg & Larkin, 1968). To do this, target stimuli were presented on the majority of trials in experimental conditions. On remaining trials, probe stimuli centred about the target stimulus were presented.

The first experiment was a replication of previous experiments using pure tones (Greenberg & Larkin, 1968) and served as a control experiment for the second experiment. The methods of experiment 1 were replicated in experiment 2; however, pure tone stimuli were replaced with sentence stimuli. In experiment 2, listeners were required to identify two key words in sentences spoken by a talker with a target  $f_0$ , and then sentences spoken by a talker with a probe  $f_0$  were inserted on a small proportion of trials. Attention functions could then be derived for each participant. This revealed the presence, or absence, of an attentional filter for  $f_0$ .

### ***2.0.1 Participants***

Twenty normal hearing participants (12 females and 8 males) between the ages of 22 and 49 years old (median = 24 years) participated in this research. Participants were recruited from within the University of Canterbury community (Christchurch, New Zealand). The University of Canterbury's Institutional Ethics Review Committee approved all experimental procedures involving participants. Prior to participating, all participants were informed of the



purpose and procedure of the study (Appendix A). Informed consent was obtained from each participant (Appendix B). All participants had received audiometric testing within the past year and were classified as having normal hearing (20 dB hearing level or lower) at octave frequencies 250 to 8000 Hz. All participants were native NZ English speakers and were compensated for their participation with a \$30.00 shopping voucher.

## **2.0.2 Instrumentation**

The computer programme MATLAB 7 (The Mathworks, Inc) was used for stimulus generation. All stimulus waveforms were sent from the built in soundcard of a Windows PC. The sound card was calibrated so that levels in dB SPL were correct to within 1 - 2 dB. To check voltage, a Tektronix TDS2002 oscilloscope was used to measure sound-card output voltage, to determine the gain that produced the intended level.

For data collection, tones and masking noises were synthesised in MATLAB 7 for the first experiment. For the second experiment, the speech stimuli were retrieved from computer files. All stimuli were presented diotically over Sennheiser HD280 Pro circumaural headphones.

Results were saved to files in MATLAB 7, and later transferred to a Microsoft Excel (2010) spreadsheet to obtain descriptive statistics. Data were later transferred to the Graphpad Prism Programme, which was used for figure generation. All data were then entered and analysed using The Statistical Package for the Social Sciences (SPSS version 20).

## **2. 1. Experiment 1**

### **2.1.1 Stimuli**

The target signal was a 1000 Hz pure tone. Probe signals were pure tones with either a 50 Hz change ( $\Delta f = 50$  Hz; condition 2), or a 100 Hz change ( $\Delta f = 100$  Hz; condition 3) above and below the target frequency. The masker was 8 kHz low-pass broadband noise. The

masker had a total duration of 440 ms in each interval. In intervals containing the signal, after 20 ms of noise alone, a 400 ms pure tone signal was presented mixed with the noise, followed by a further 20 ms of noise alone. The interstimulus interval was 100ms. All tones were presented at the same signal-to-noise ratio (SNR) throughout the conditions. The masking noise was fixed and had a constant overall level of 65 dB SPL, and the signal level was chosen by the experimenter for each participant (described later).

### 2.1.2 Procedure

Participants were seated in front of a lap top screen. This screen displayed two boxes, one for interval 1 (labelled “Int1”) and one for interval 2 (labelled “Int2”; see Figure 8). For each trial, participants heard two successive intervals each containing noise, and only one interval containing the signal. The order of sounds was random. The participants’ task was to report which interval contained the stimulus of interest. Participants were instructed to indicate their response by clicking on the interval with a tonal signal present (for a full account of participant instructions, see Appendix C). If subjects were unsure about which interval contained the signal, they were encouraged to guess in order to continue in the trial block. After the response was recorded, one box would briefly flash white to indicate which interval was correct.



**Figure 8.** A screenshot of the computer screen display response matrix for Experiment 1. The participant selected the interval which they thought contained a tone

Participants first completed a practice condition. This allowed participants to familiarise themselves with the task before data collection began. This was a single-

frequency condition using only 1000 Hz signals which matched the target signal for subsequent experimental conditions in every respect (except level) to familiarise listeners with this tone for detection. A block of 12 trials at a SNR which made the pure-tone signal highly detectable was presented first (-10 dB SNR). No data were saved from this condition; however the proportion of correct responses was displayed at the end for the experimenter to monitor participant performance. Following the first block at an SNR of -10 dB, further practice blocks were presented with an SNR which was reduced so that the proportion of responses obtained fell between 70 and 90% correct. This would reflect a performance which is better than chance (50%), but less than perfect. If the chosen SNR yielded a performance outside this range, the SNR was adjusted, and the practice condition was repeated. This process was repeated until detection rates fell within the desired range. The SNR remained fixed throughout the following conditions.

All participants completed all conditions once. Each condition contained blocks of 168 trials each. For experimental conditions, 120 trials (72%) contained target stimuli, and 48 trials (28%) contained probe stimuli. In each experimental condition, a maximum of two probe frequencies were presented. One was above the target stimulus, whereas the other was the same distance below the target stimulus. In control conditions, all 168 trials contained target stimuli. Table 1 lists each of the conditions, including the characteristics of the different stimuli presented, and the number of trials dedicated to each. The order of trials was determined randomly, with the exception that no probe stimuli were allowed to occur in the first 10 trials of each block.

Participants had unlimited response time on each trial. A prompt would appear every 25 trials offering the participants an opportunity for a rest period, which participants were encouraged to utilise as needed. The participant could select a box labelled “click to start/resume” to continue the block. The order of the conditions was determined using a

random number generator. Each block lasted approximately 6-7 minutes, depending on the participant's response time and number of breaks taken.

Table 1.  
Characteristics of Conditions for Experiment 1: Pure Tones

| Condition  | Stimuli                                   | Number of Trials   |
|--|---|--|
| Practice   | Target = 1000 Hz                          | 12 target trials<br>(multiple blocks)  |
| 1. Target-only (control)                         | Target = 1000 Hz                          | 168 target trials  |
| 2. Target $\pm$ 50 Hz<br>( $\Delta f$ = 50 Hz)   | Target = 1000 Hz<br>Probes = 950, 1050 Hz | 120 target trials<br>24 probe trials at 950 Hz<br>24 probe trials at 1050 Hz |
| 3. Target $\pm$ 100 Hz<br>( $\Delta f$ = 100 Hz) | Target = 1000 Hz<br>Probe = 900, 1100 Hz  | 120 target trials<br>24 probe trials at 900 Hz<br>24 probe trials at 1100 Hz |
| 4. Target-only (control)                         | Target = 900 Hz                           | 168 target trials  |
| 5. Target-only (control)                         | Target = 1100 Hz                          | 168 target trials  |

## 2.2 *Experiment 2*

### 2.2.1 Stimuli

The second experiment used sentence-length speech materials taken from the Coordinate Response Measure (CRM) corpus (Bolia, Nelson, Ericson & Simpson, 2000). The CRM corpus has gained wide acceptance as a research tool for investigating speech intelligibility in background noise (Brungart, 2001; Eddins & Liu, 2012). It employs a closed-set task which requires participants to listen to sentences with two target words (discussed next). The same sentence can be presented repeatedly without concern about effects of set size or variation in difficulty, and so these are appropriate stimuli to minimize possible effects of practice. Furthermore, the corpus has the advantage of carrying no linguistic

context, and therefore reduces effects of memory and language abilities. The corpus contains a large number of speech samples from both female and male talkers. Each sentence in the complete corpus is spoken in American English by 4 male talkers and 4 female talkers with varying f0s. In these experiments, a subset of sentences produced by female ‘talker 6’ was chosen for use.

### **2.2.2 Signal Manipulation**

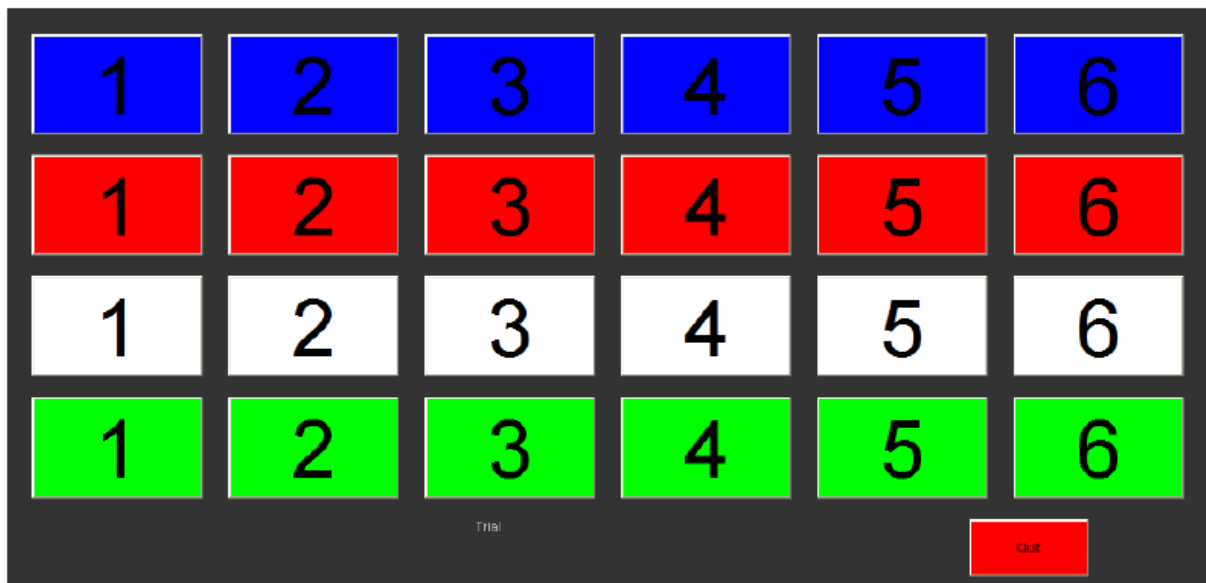
Initially, each sentence contained three key words embedded in the carrier phrase “Ready (call sign) go to (colour) (number) now.” The “Ready (call sign)” portion was removed to make each sentence shorter and ultimately reduce test time. Sentences in the complete CRM corpus comprise 4 colours (“blue,” “green,” “red,” “white,”), and 8 possible numbers (1 through 8). For this experiment, the numbers 7 and 8 were removed from the set to exclude the bisyllabic number 7. This reduced the number of unique sentences, and also resulted in a more practical ratio of probe signals to target signals. The remaining set contained 24 possible sentence combinations of four colours and six numbers.

The sentences were processed to flatten the contour of f0 of the female talker in order to achieve the unvarying target frequency of 220 Hz for the chosen female talker. This processing was done using Praat speech analysis software (Phonetic Sciences, Amsterdam). Probe sentences were also created by editing each manipulated target sentence. Each target sentence was edited four times. This created four sets of 24 probe sentences. The first and second sets of probes had unvarying f0s of either plus or minus 5 Hz ( $\Delta f_0 = 5$  Hz), whereas the third and fourth sets had unvarying f0s of either plus or minus 10 Hz ( $\Delta f_0 = 10$  Hz), with respect to the target f0. For a full account of the manipulation procedure see Appendix D. These manipulations give the sentences a monotone characteristic, reducing prosodic cues such as intonation and stress; however it should not affect intelligibility as compared to the original sentences.

The sentences were saved in computer-readable files and then mixed with a speech-shaped masking noise. This noise spectrum matched the long-term spectrum of the all CRM sentences spoken by female talkers. The overall level of the masking noise was fixed and presented at an overall level of 65 dB SPL. The level of the speech stimuli was chosen by the experimenter for each participant (discussed next), and remained fixed throughout the different conditions.

### **2.2.3 Procedure**

Procedures were essentially the same as those for experiment 1, however, sentence materials replaced pure tones. Before beginning testing, participants were given verbal instructions for completing the experiment (for a summary, see Appendix C). Participants were seated in front of a lap top screen displaying a four-row, six-column matrix of response buttons with colours and numbers that matched the key words in the CRM sentences (see Figure 9). On each trial, the participant's task was to identify the colour and number heard by using the computer mouse to select the appropriate response button. Feedback was provided on every trial. After the participant made a selection, one box would briefly change colour to indicate which colour-number pair was correct.



**Figure 9.** A screenshot of the computer screen display response matrix for Experiment 2. The participant selected the colour-number pair they thought they heard

The participants first completed a 12-trial practice condition containing only sentences at the target f0. For the first block of practice trials, a highly identifiable SNR of -5.0 dB was chosen. This allowed participants to become familiar with the listening task and with the f0 of the target sentences. A response was scored as correct if both the colour and the number were identified. The SNR was adjusted on subsequent blocks so that the proportion of correct responses obtained fell between 40 and 60%, well above the performance attainable by guessing (which is approximately 4%). Once this performance level was reached, a final block at the same SNR was completed to check that performance was relatively consistent.

Each condition lasted approximately 7-8 minutes, depending on the participant's response time and number of breaks taken. A break of about 1-2 minutes was provided between each block, while waveforms were loaded for the following block. A standard session in Experiment 2 consisted of the practice condition and three experimental

conditions. The characteristics of the stimuli, and number of trials containing each stimulus within each condition are listed in Table 2.

Table 2.

Characteristics of Conditions for Experiment 2: Sentence Identification

| Condition  | Stimuli  | Number of Trials  |
|--|--|---|
| Practice (targets)   | Target $f_0 = 220\text{Hz}$                                    | 12 target trials<br>(multiple blocks)                                       |
| 1. Target-only   | Target $f_0 = 220\text{ Hz}$                                   | 168 target trials   |
| 2. Target $\pm 5\text{ Hz}$ probes<br>( $\Delta f_0 = 5\text{ Hz}$ )   | Target $f_0 = 220\text{ Hz}$<br>Probe $f_0$ s = 215 Hz, 225 Hz | 120 target trials<br>24 probe trials at 215 Hz<br>24 probe trials at 225 Hz |
| 3. Target $\pm 10\text{ Hz}$ probes<br>( $\Delta f_0 = 10\text{ Hz}$ ) | Target $f_0 = 220\text{ Hz}$<br>Probe $f_0$ s = 210 Hz, 230 Hz | 120 target trials<br>24 probe trials at 210 Hz<br>24 probe trials at 230 Hz |

### 2.3 Data Analysis

A series of one-way repeated measures analyses of variances (ANOVAs) in SPSS (version 20) were used for the first experiment to determine the effect of frequency on the ability to correctly identify which interval contained the tone. The same series of statistical tests was used in the second experiment to determine the effect of  $f_0$  on the ability to identify target words in the sentence identification task. Pairwise comparisons were analysed using the Bonferroni method. A  $p$  value of  $< .05$  was deemed to be significant for both experiments.



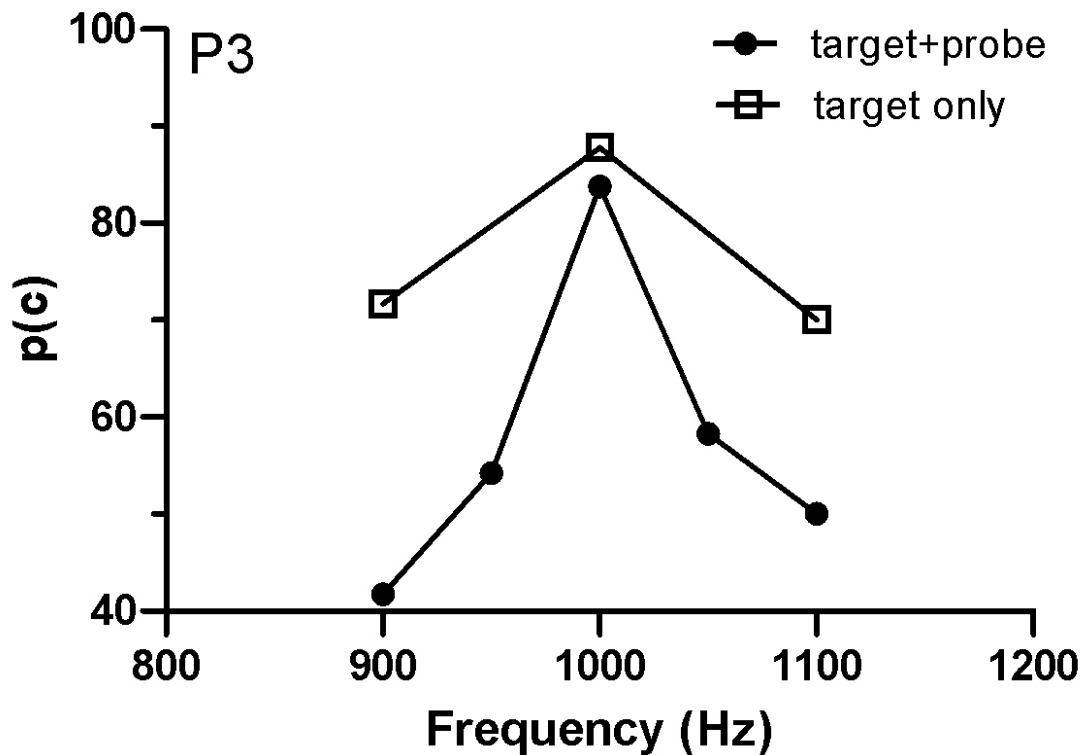
## Chapter Three: Results

---

### 3.1 *Experiment 1*

All 20 participants were easily able to perform the tone detection task at a SNR of -10 dB in the first block of practice trials. The SNR was then adjusted as necessary for each individual so that their rate of correct tone detection hovered about 80% for the target frequency signal. The range of SNRs for participants to achieve these detection rates was between -20 and -24 dB SNR.

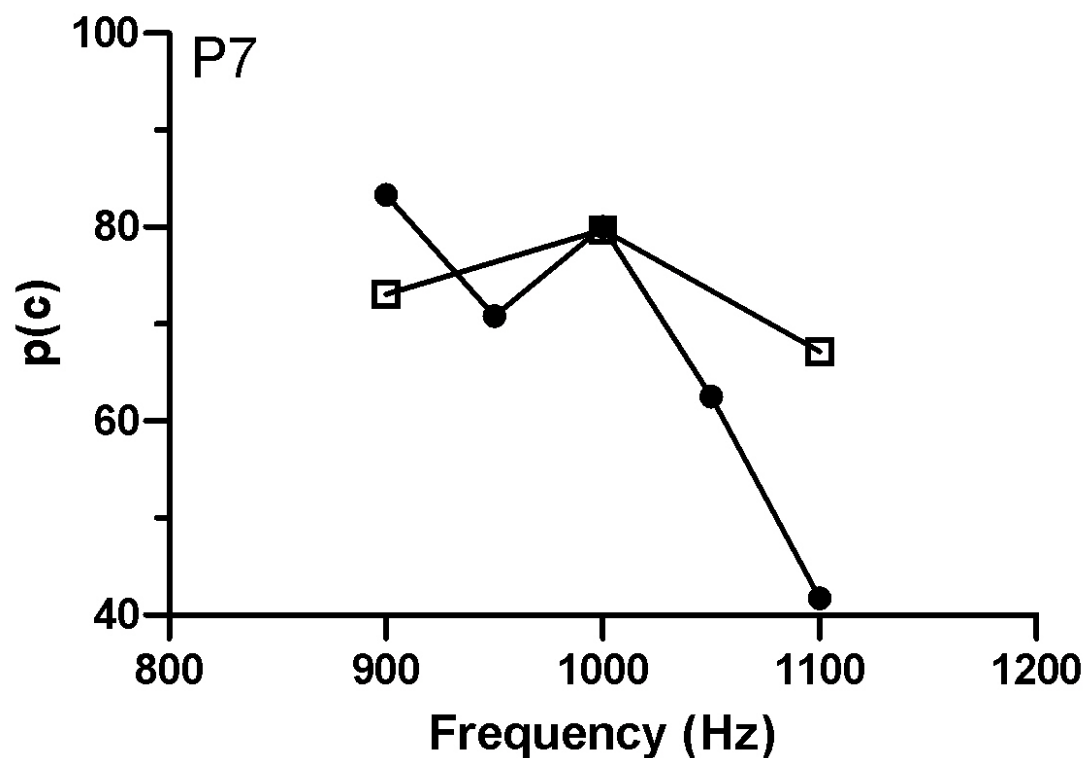
The percentage of correct responses (percent correct) in detecting the target tone varied across presentation conditions as well as signal frequency. Various patterns of performance could be observed. An example of the most prevalent pattern, as illustrated with the results from Participant 3, is shown in Figure 10. In conditions where targets and probes were presented (T+P conditions), the percent correct scores in detecting the target frequency at 1000 Hz were the highest at 84%. When  $\Delta f = 50$  Hz, percent correct scores decreased with respect to the target frequency. When  $\Delta f = 100$  Hz, performance had declined further to chance levels (50% or less). In conditions where the 900 and 1100 Hz probe signal frequencies were presented alone as targets (T-only conditions), percent correct scores were markedly improved, compared to those for the same signals in T+P conditions. The findings that the performance of tone detection was better (1) when probe tones were presented alone as targets, than when they acted as probe tones; and (2) with probe tones closer in frequencies to the target tone than with probe tones at frequencies further away from the target tone; are consistent with those of previous studies (Greenberg & Larkin, 1968; Scharf et al., 1987; Dai et al., 1991). However, the performance in the T-only conditions was found in this study to be better for 1000 Hz signals than for 900 and 1100 Hz signals. In previous studies they found consistent performances across frequencies in T-only conditions (Greenberg & Larkin, 1968; Scarf et al., 1987; Dai et al., 1991).



**Figure 10.** Percent correct scores,  $p(c)$ , for Participant 3 in Experiment 1 (tone detection task). The open squares represent the conditions where only one frequency was presented. The closed circles represent the results from conditions in which both target and probe signals were presented. The data points for the T-only conditions are based on 168 trials per point. The data points for probe frequencies (900, 950, 1050 and 1100 Hz) in the T+P conditions are based on 24 trials per point. To simplify presentation of the results, the data point for the target in the two T+P conditions is the average of the performance for this signal in these conditions, and therefore this data point is based on 240 trials.

Some participants did not show a clear pattern in their responses. Others demonstrated better detection of probe tones with a  $\Delta f$  of 100 Hz either above or below the target signal. An example of a participant who demonstrated better detection of the 900 Hz probe tone, as compared to the 1100 Hz probe tone is shown in Figure 11. This participant shows a pattern of the percent correct scores in detecting the target or the higher frequency probe tones (1050 and 1100 Hz) decreasing as the  $\Delta f$  increases, which is consistent with the presence of an

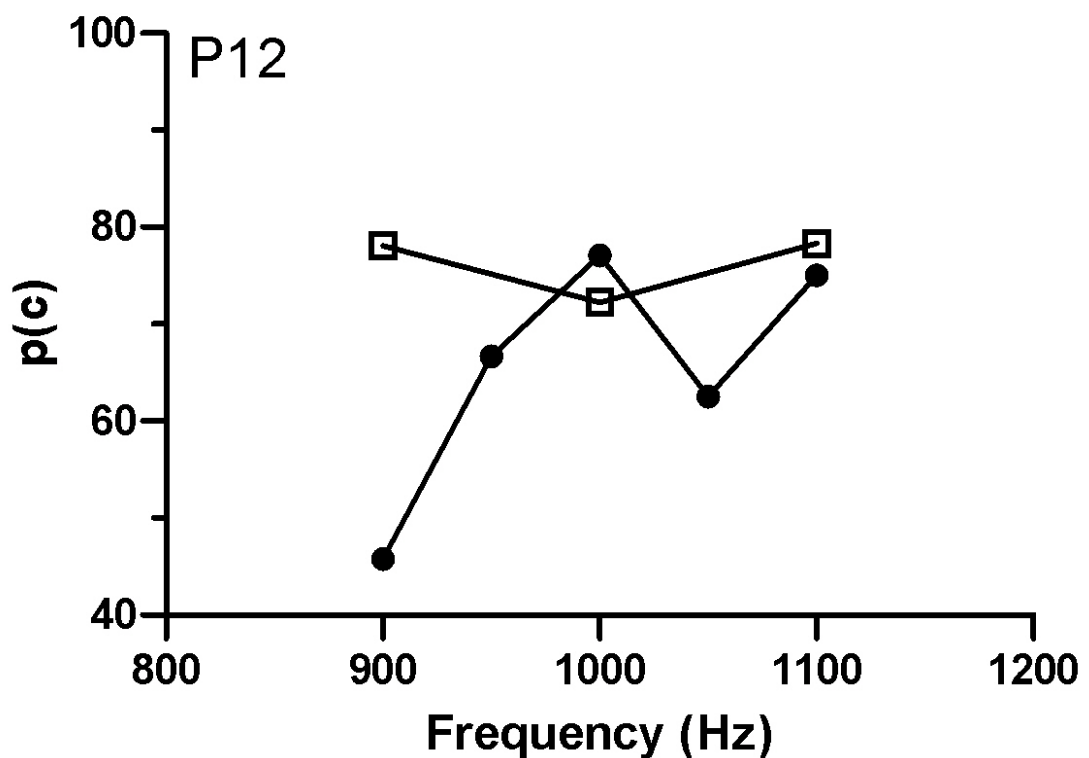
attentional filter for frequency. A similar pattern was found for the 950 Hz probe; however, detection of the 900 Hz probe tones was higher than that which would be expected from an attentional filter mechanism (Greenberg & Larkin, 1968; Scharf et al., 1987). The average detection for the 900 Hz probe tone was 83%, which is almost equal to the average detection of the target (80%) in the T+P conditions. With only a small difference between the percent correct scores for the three target frequencies (smaller than 15%), performance in the T-only conditions appeared to be relatively stable across frequencies.



**Figure 11.** Percent correct scores,  $p(c)$ , for Participant 7 in Experiment 1 (tone detection task). Symbols represent the same conditions as for Participant 3 (Fig 10) and are based on the same number of trials per point.

In contrast, an example of a participant who demonstrated poorer tone detection with the 900 Hz probe tone than with the 1100 Hz probe tone is shown in Figure 12. Consistent with other participants, the detection of the target signal in the T+P condition was about 80%.

The percent correct scores in tone detection with probe tones below the target signal declined as they deviated from the target frequency. The percent correct scores for detecting the 1050 Hz probe tone is similar to that for the 950 Hz probe tone. The finding that performance for probes with a  $\Delta f$  of 50 Hz is poorer than that for the target is consistent with what might be expected based on previous studies (Greenberg & Larkin; Scharf et al., 1987). Detection of the 1100 Hz probe, however, is approximately equal to that of the target signal, contrary to previous research. Performance in the T-only conditions was consistent across the frequency range tested.

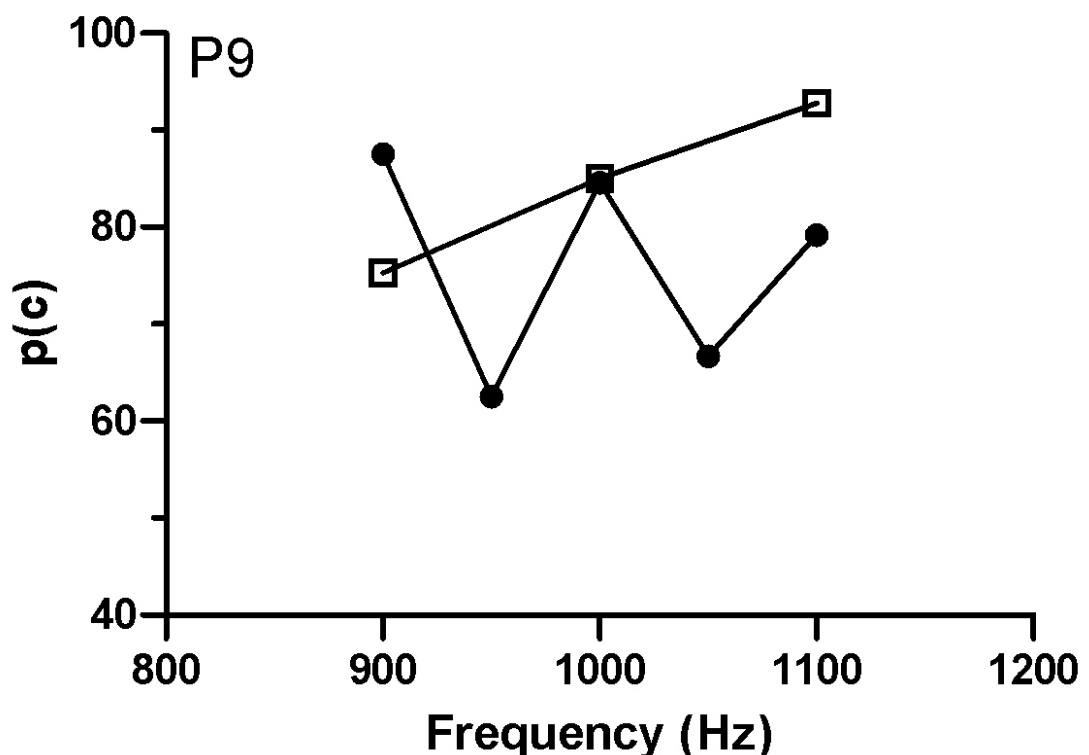


**Figure 12.** Percent correct scores,  $p(c)$ , for Participant 12 in Experiment 1 (tone detection task). Symbols represent the same conditions as for participant 3 (Fig 10) and are based on the same number of trials per point.

Although most of participants performed worse in the T+P condition as  $\Delta f$  increased, some participants performed better when probes had a  $\Delta f$  of 100 Hz than for detection of probe tones with a  $\Delta f$  of 50 Hz. For example, the performance of Participant 9 hovered about

80 – 90% for the target and  $\Delta f = 100$  Hz probe signals (see Figure 13). When  $\Delta f = 50$  Hz, however, the percent correct score dropped to between 60 – 65%. In the T-only conditions, performance improved as the frequency increased, from 75% for 900 Hz probes, to 93% for 1100 Hz probes.

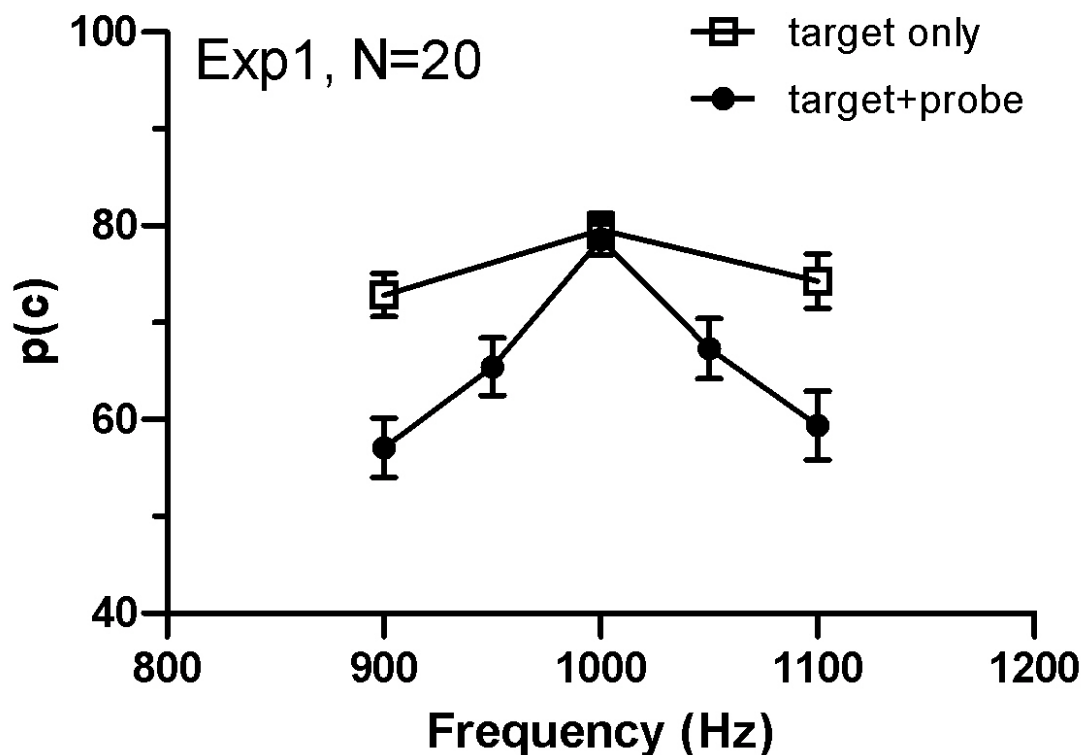
Although there was a large range of different response patterns across participants, some participants' attention functions exhibited a good approximation of the expected attentional filter tuned to the target frequency (Greenberg & Larkin, 1968; Scharf et al., 1987). Despite some individuals showing better detection of certain probe frequencies with respect to other probe frequencies, there was only one instance in which the probe signal was detected better than the target signal in a T+P condition (see Figure 11; detection of the 900 Hz probe).



**Figure 13.** Percent correct scores,  $p(c)$ , for Participant 9 in Experiment 1 (tone detection task). Symbols represent the same conditions as for participant 3 (Fig 10) and are based on the same number of trials per point.

When assessing the data on a group basis, there is a clear demonstration of an effect of frequency on tone detection performance. When the frequency of the tone matched the anticipated frequency (target trials), the overall detection rate was higher than it was when the tone was an off-frequency probe tone. Figure 14 shows the mean percentage scores as a function of frequency.

The mean detection of probes in T+P conditions was poorer than detection of the same tones in conditions where they were presented alone as target tones. There was a greater reduction in tone detection in the T+P conditions for probes further in frequency from the target than for those nearer in frequency. The overall pattern of results revealed that the participants exhibited the expected pattern of responses for the probe-signal method (Greenberg & Larkin, 1968; Scharf et al., 1987), consistent with the presence of an attentional filter for frequency.



**Figure 14.** Mean Percent correct scores,  $p(c)$ , for all participants in Experiment 1 (tone detection task). Each data point represents the mean across all

participants for each frequency. The open squares represent the conditions where only one frequency was presented. The closed circles represent the results from conditions in which both target and probe signals were presented. To simplify presentation of the results, the data point for the target in the two T+P conditions is the average of the performance for this signal in these conditions. Error bars are  $\pm 1$  standard error.

### 3.1.2 Statistical Analysis for Experiment 1

Descriptive statistics for the mean data are presented in Table 3. A series of one-way repeated measures (RM) Analysis of Variances (ANOVAs) were conducted on individuals' percent correct scores in tone detection for each presentation condition separately, with stimulus frequency as the independent variable. Signal frequency consisted of three levels in the T+P condition having probes with a  $\Delta f$  of 50 Hz: 950, 1000 and 1050 Hz, three levels in the T+P condition having probes with a  $\Delta f$  of 100 Hz: 900, 1000 and 1100 Hz, and three levels in the T-only conditions: 900, 1000 and 1100 Hz.

The first one-way RM ANOVA was run on the individuals' percent correct scores in the T+P condition containing probes with a  $\Delta f$  of 50 Hz. The assumptions of RM ANOVA were tested. Mauchly's test indicated that the assumption of sphericity had been violated,  $\chi^2(2) = 9.724, p = 0.008$ ; therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ( $\epsilon = 0.71$ ). Normality of residuals was verified visually with the help of Q-Q plots. Results from the RM ANOVA revealed a statistically significant effect of stimulus frequency in the T+P condition using a  $\Delta f$  of 50 Hz [ $F(1.41, 26.81) = 12.16, p < 0.001$ ]. Pairwise comparisons using the Bonferroni adjustment method showed that participants detected the 1000 Hz target tone with a significantly higher mean percentage correct score than that obtained for both the 950 Hz ( $p < 0.001$ ) and 1050 Hz ( $p < .001$ ) probes.

The second one-way RM ANOVA was run on the individuals' percent correct scores in the T+P condition containing probes with a  $\Delta f$  of 100 Hz. Mauchly's test indicated that the assumption of sphericity had been met,  $\chi^2(2) = 2.80, p = 0.246$ . Results from the RM

ANOVA revealed a statistically significant effect of stimulus frequency in the T+P condition using a  $\Delta f$  of 100 Hz [ $F(2, 38) = 17.61, p < 0.001$ ]. Pairwise comparisons using the Bonferroni adjustment method showed that participants detected the 1000 Hz target with a significantly higher percentage correct scores than the 900 Hz ( $p < 0.001$ ) and 1100 Hz probes ( $p < 0.001$ ).

When presented alone as targets (Conditions 1, 4 and 5), the 900, 1000, and 1100 Hz signals were correctly detected on the majority of trials ( $> 70\%$ , see Table 3). A one-way RM ANOVA was run on the individuals' percent correct scores in the T-only condition. Mauchly's test indicated that the assumption of sphericity had been met,  $\chi^2(2) = 0.312, p = 0.856$ . As expected, results from the RM ANOVA revealed no statistically significant effect of stimulus frequency in the T-only condition [ $F(2, 38) = 3.296, p = 0.085$ ].

Additional pairwise comparison procedures were conducted on the individuals' percent correct scores for the different frequency tones used as probes or targets across all conditions. As expected (Greenberg & Larkin, 1968; Scharf et al., 1987), there was a systematic decline in the detection of probe tones as they deviated from the target signal. The 950 Hz probe showed a higher percentage correct score than the 900 Hz probe, and the 1050 Hz probe had a higher percent correct score than the 1100 Hz probe. A pairwise comparison of these results (900 Hz versus 950 Hz; 1050 Hz versus 1100 Hz), however, did not reach statistical significance (*SPSS Bonferroni adjusted*,  $p = 1.00$ ). A comparison was also made between the individuals' percent scores in detecting the 900 Hz and 1100 Hz signals when they were presented in either T+P or T-only conditions. The results from pairwise comparisons revealed a significant difference between the detectability of the 900 Hz signal when it acted as a probe signal, and the detectability of the same signal when it acted as a target signal ( $p < 0.05$ ). The same significant difference was found between the mean percent correct scores for detecting the 1100 Hz signal in the T+P and T-only conditions ( $p < 0.05$ ).



Table 3. Descriptive statistics for the five conditions for tone detection

| Condition  | Stimuli |           | Statistics     |       |
|--|---------|-----------|----------------|-------|
|  | Type    | Frequency | Mean % correct | SD    |
| 1. Target-only                                   | Target  | 1000 Hz   | 79.55%         | 9.61  |
| 2. Target $\pm$ 50 Hz<br>( $\Delta f = 50$ Hz)   | Probe   | 950 Hz    | 65.43%         | 13.29 |
|  | Target  | 1000 Hz   | 79.58%         | 7.16  |
|  | Probe   | 1050 Hz   | 67.30%         | 13.87 |
| 3. Target $\pm$ 100 Hz<br>( $\Delta f = 100$ Hz) | Probe   | 900 Hz    | 57.08%         | 13.59 |
|  | Target  | 1000 Hz   | 77.63%         | 7.06  |
|  | Probe   | 1100 Hz   | 59.38%         | 15.82 |
| 4. Target-only                                   | Target  | 900 Hz    | 72.83%         | 11.81 |
| 5. Target-only                                   | Target  | 1100 Hz   | 74.26%         | 13.55 |

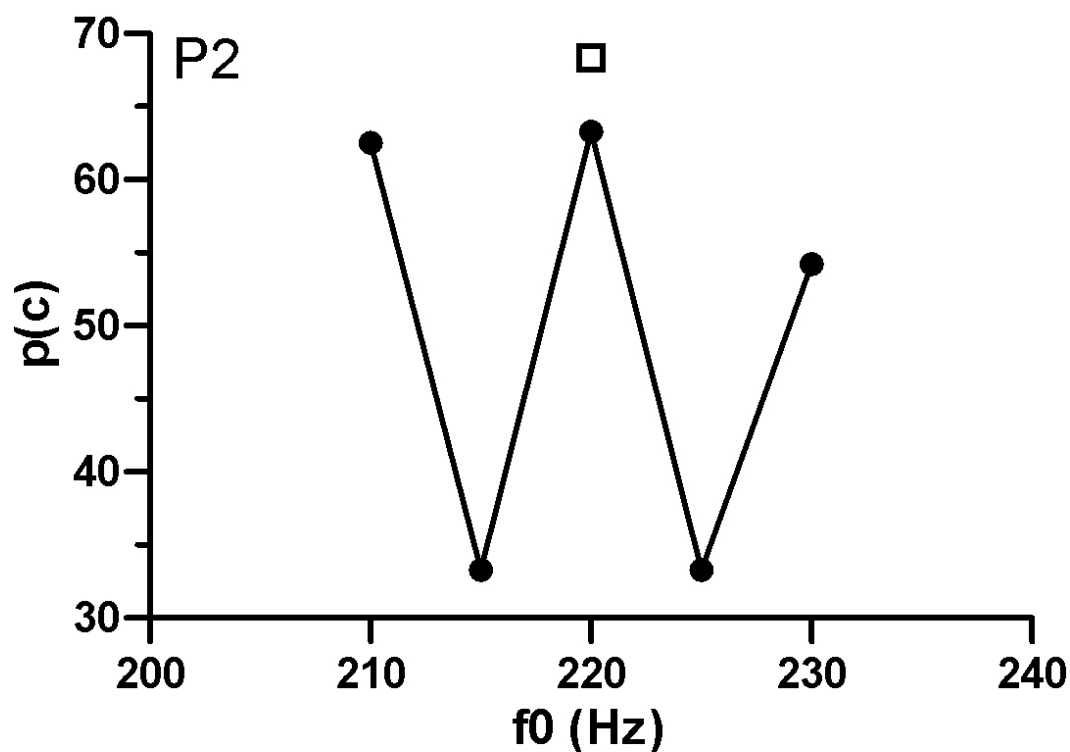
*Note.* For each condition in Experiment 1, the type of stimulus (target or probe) and the frequency of the signal are listed. The final two columns list the mean percentage correct and standard deviations (SD) for the target and probe signals in each condition.

### 3.2 Experiment 2

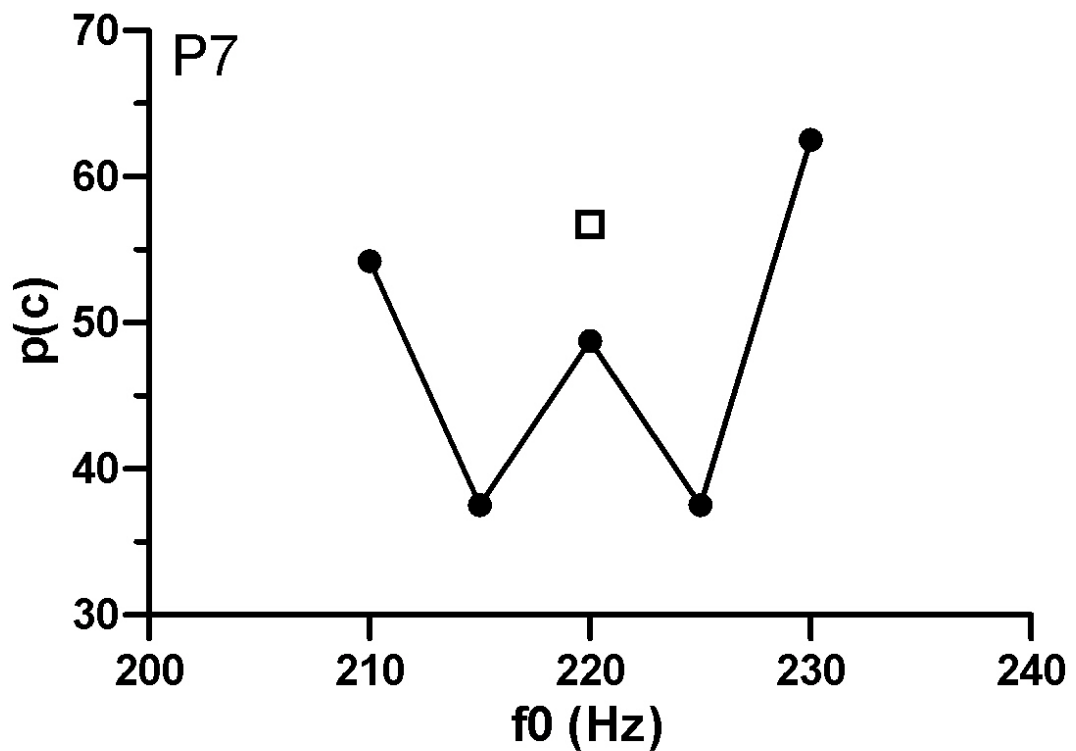
All participants were able to complete the sentence identification task with little practice. The practice condition was repeated while the SNR was adjusted for each participant so that their rate of correct tone detection hovered about 50% for the correct identification of colour-number pairs in target f0 sentences. Signal to noise ratios fell within the range of -8 to -10 dB across participants.

Similar to Experiment 1, there was a large amount of variability in performance within and across the conditions. To illustrate the range of response patterns, a series of individual results from a selection of participants have been chosen. The results of four participants were selected to illustrate two distinct patterns in the results.

The first pattern, which is displayed in Figures 15 and 16, consists of better detection of the target and probes with a  $\Delta f_0$  of 10 Hz, compared to probes with a  $\Delta f_0$  of 5 Hz. For participant 2 (see Figure 15) the performance for the target and probes with a  $\Delta f_0$  of 10 Hz ranged between 54 and 63%, whereas the performance for sentences with a  $\Delta f_0$  of 5 Hz was 33%. Similarly, Participant 7's performance on probes with a  $\Delta f_0$  of 10 Hz ranged from 54 – 63%, and correct identification of the target was 48% (see Figure 16). Performance on the probe sentences with a  $\Delta f$  of 5 Hz, however, was considerably lower at 38%.

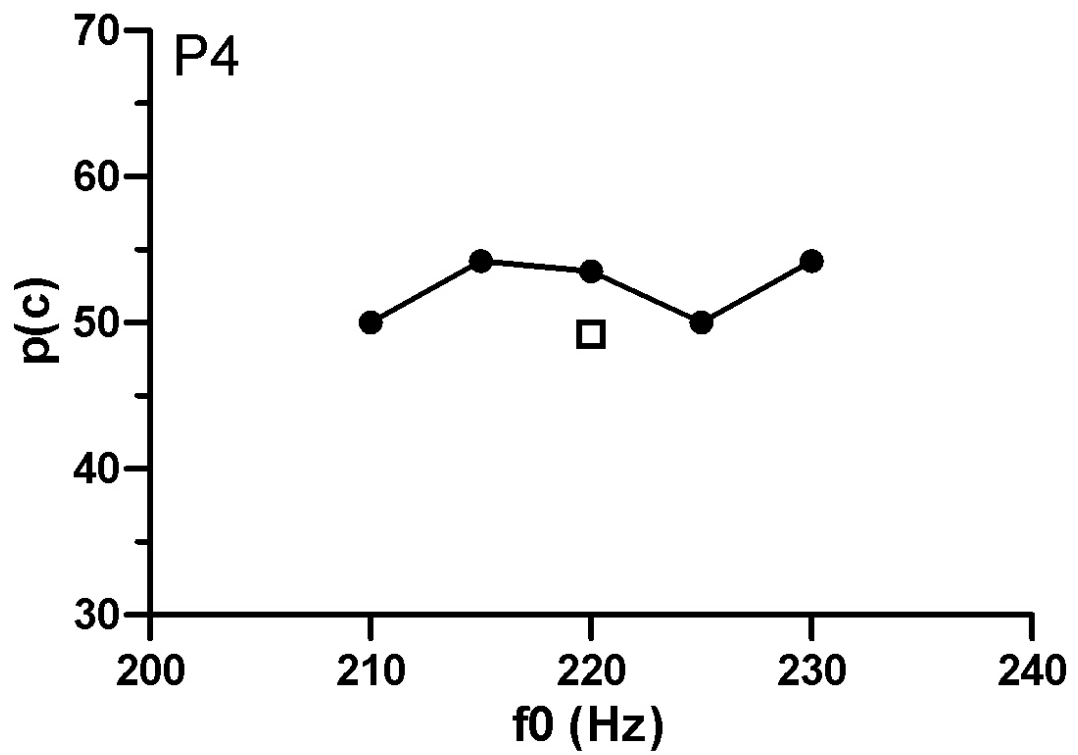


**Figure 15.** Percent correct scores,  $p(c)$ , for Participant 2 in Experiment 2 (sentence identification task). The open square represents the condition where sentences of only one  $f_0$  were presented. The closed circles represent the results from conditions in which both target and probe sentences were presented. The data point for the T-only conditions is based on 168 trials per point. The data points for probe  $f_0$ s (210, 215, 225 and 230 Hz) in the T+P conditions are based on 24 trials per point. To simplify presentation of the results, the data point for the target in the two T+P conditions is the average of the performance for this signal in these conditions, and therefore this data point is based on 240 trials.

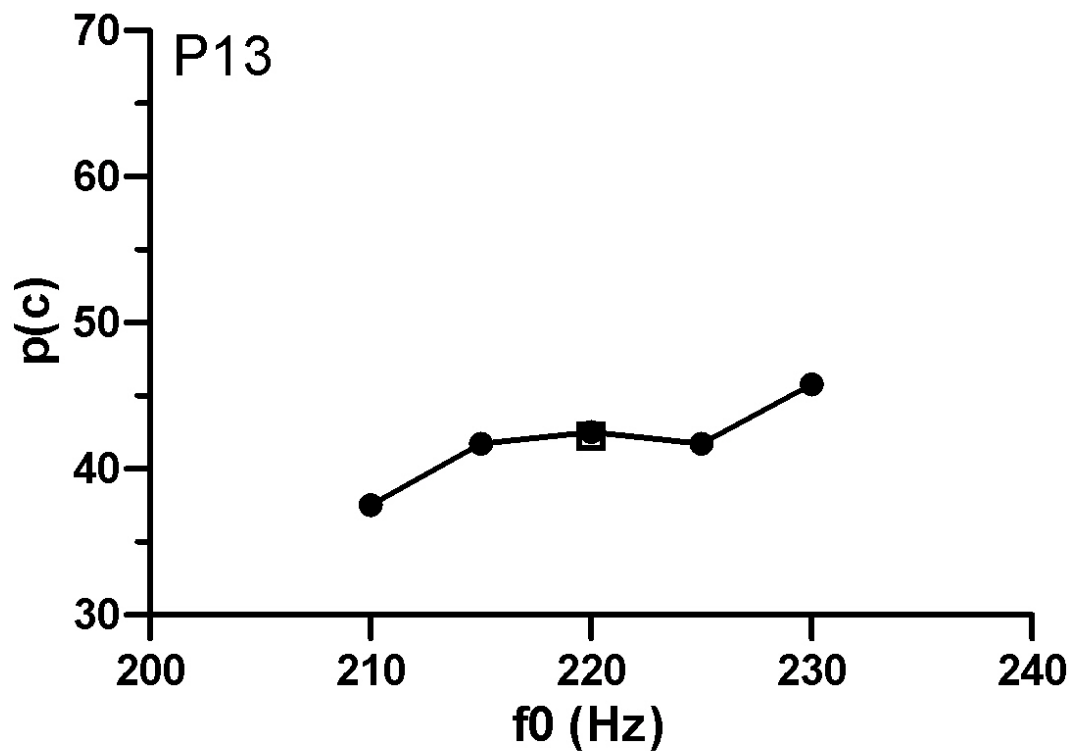


**Figure 16.** Percent correct scores,  $p(c)$ , for Participant 7 in Experiment 2. Symbols represent the same conditions as for participant 2 (Fig 15) and are based on the same number of trials per point.

Other participants showed response patterns that were independent of  $f_0$ . The results from two of these participants are displayed in Figures 17 and 18. The results from both of these participants show very little variation in performance as a function of  $\Delta f_0$ . Performance ranged between 50 - 54 % for participant 4 and 38 – 46% for participant 13, across all conditions.

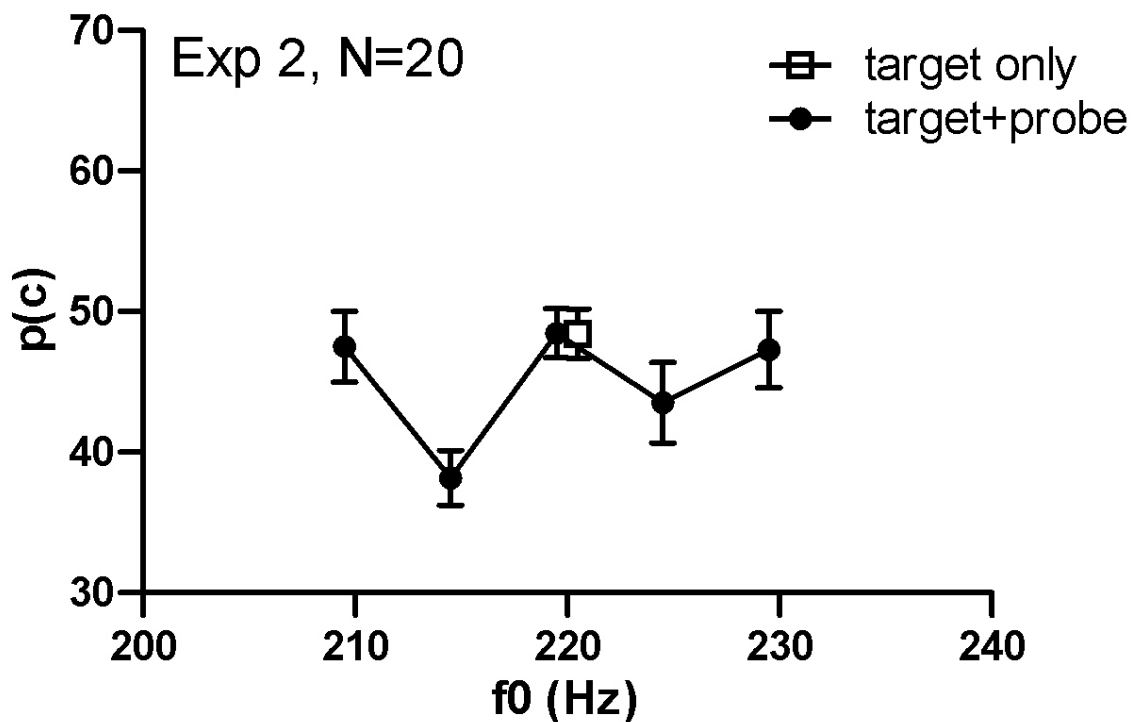


**Figure 17.** Percent correct scores,  $p(c)$ , for Participant 4 in Experiment 2. Symbols represent the same conditions as for participant 2 (Fig 15) and are based on the same number of trials per point.



**Figure 18.** Percent correct scores,  $p(c)$ , for Participant 13 in Experiment 2. Symbols represent the same conditions as for participant 2 (Fig 15) and are based on the same number of trials per point.

Mean percent correct scores for the sentence identification task are presented for each  $f_0$  in Figure 19. As shown in Figure 10, percent correct scores hovered around 50% for the 220 Hz sentences regardless of whether the sentence was presented in a T-only condition or a T+P condition. Performance was similar for probe sentences with a  $\Delta f_0$  of 10 Hz (47 - 48%). For probe sentences with a  $\Delta f_0$  of 5 Hz, however, the overall performance declined to 38% for the 215 Hz probe, and 43% for the 225 Hz probe.



**Figure 19.** Mean Percent correct scores,  $p(c)$ , for all participants in Experiment 2 (sentence identification task) as a function of  $f_0$ . To simplify presentation of the results, the data point for the target in the two T+P conditions is the average of the performance for this signal in these conditions. The data point for the T-only condition at 220 Hz was displaced horizontally to more clearly demonstrate the results as compared to the data point at 220 Hz for the T+P condition. Error bars are  $\pm 1$  standard error.

### 3.2.1 Statistical Analysis for Experiment 2

Table 4 displays the descriptive statistics for the group data in Experiment 2. A series of one-way RM ANOVAs were conducted on the percent correct scores for identification of correct colour-number pairs. The independent variable, or within subjects factor, for the ANOVAs conducted for the two T+P conditions separately was stimulus  $f_0$ . For the T+P condition having probes with a  $\Delta f_0$  of 5 Hz, there were three  $f_0$  levels: 215, 220, and 225 Hz. For the T+P condition having probes with a  $\Delta f_0$  of 10 Hz, there were three levels: 210, 220 and 230 Hz. The independent variable for the ANOVA conducted for data related to the target  $f_0$  of 220 Hz was presentation condition, which consisted of three levels: T-only,

Target  $\pm 5$  Hz probes, and Target  $\pm 10$  Hz probes.

A one-way RM ANOVA was conducted on the participants' percent correct scores in the T+P condition containing probes with a  $\Delta f_0$  of 5 Hz. Mauchly's test indicated that the assumption of sphericity had been violated,  $\chi^2(2) = 9.724$ ,  $p = 0.008$ ; therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ( $\epsilon = 0.77$ ). Normality of residuals was verified visually with the help of Q-Q plots. The one-way RM ANOVA results revealed a significant effect of stimulus  $f_0$  on the percent correct scores [ $F(1.54, 29.26) = 10.84$ ,  $p = 0.001$ ]. Pairwise comparisons using the Bonferroni adjustment method revealed that individuals' identification of the 215 Hz probe sentences was significantly poorer than that for the 220 Hz target sentences ( $p < 0.001$ ). Although the mean identification of the 225 Hz probe sentences was poorer than that of target sentence in the same condition, this did not reach statistical significance ( $p = 0.08$ ).

Another one-way RM ANOVA was conducted on the individuals' percent correct scores in the T+P condition containing probes with a  $\Delta f_0$  of 10 Hz. Mauchly's test indicated that the assumption of sphericity had been violated,  $\chi^2(2) = 10.56$ ,  $p = 0.005$ ; therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ( $\epsilon = 0.693$ ). Normality of residuals was verified visually with the help of Q-Q plots. Results from the one-way RM ANOVA revealed that there was no significant stimulus  $f_0$  effect on the identification performance [ $F(1.39, 26.32) = 0.117$ ,  $p > 0.05$ ].

The percent correct scores for identifying speech at a target  $f_0$  of 220 Hz across presentation conditions (i.e., T-only, Target  $\pm 5$  Hz probes, and Target  $\pm 10$  Hz probes) were submitted to a one-way RM ANOVA to determine whether there was a condition effect. Mauchly's Test of Sphericity indicated that the assumption of sphericity had been met,  $\chi^2(2) = 238$ ,  $p = 0.888$ . As expected, the results of a one-way RM ANOVA showed no significant

condition effect [ $F(2,38) = 0.009$ ,  $p = 0.991$ ] on the percent correct scores for identification of colour-number pairs in sentences with the same target  $f_0$  (220 Hz).

Table 4. Descriptive statistics for CRM sentence identification

| Condition   | Stimuli |        | Statistics     |       |
|---|---------|--------|----------------|-------|
|   | Type    | $f_0$  | Mean % correct | SD    |
| 1. Target only  | Target  | 220 Hz | 48.42%         | 9.06  |
| 2. Target $\pm$ 5 Hz probes<br>( $\Delta f_0 = 5$ Hz)   | Probe   | 215 Hz | 38.13%         | 8.70  |
|   | Target  | 220 Hz | 48.38%         | 8.05  |
|   | Probe   | 225 Hz | 43.53%         | 12.86 |
| 3. Target $\pm$ 10 Hz probes<br>( $\Delta f_0 = 10$ Hz) | Probe   | 210 Hz | 47.50%         | 11.26 |
|   | Target  | 220 Hz | 48.55%         | 7.66  |
|   | Probe   | 230 Hz | 47.29%         | 12.12 |

*Note.* For each condition in Experiment 2, the type of stimulus (target or probe) and the  $f_0$  of the sentences are listed. The final two columns list the mean percentages of correctly identified colour-number pairs and standard deviations (SD) for the target and probe sentences in each condition.



## Chapter Four: Discussion

---

The present study investigated whether there is an attentional filter in the f0 domain. This section will discuss the results of the study in relation to the research hypotheses. The implications of the data are considered with respect to other relevant studies on auditory attention, and filter systems. Limitations and possible directions for future research will be discussed throughout.

### **4.1 Experiment 1**

Experiment 1 was a control condition for Experiment 2. It was conducted to investigate the adequacy of the chosen probe-signal method to reveal an attentional filter mechanism in the frequency domain. The methods were then replicated in the second experiment, however, instead of changing the frequency of tonal signals, the f0 of sentence stimuli was manipulated.

#### **4.1.1 Experimental Findings**

The group results from experiment 1 revealed a difference in detection rates as a function of signal frequency. Mean detection performance was best for the target frequency with about 80% correct, and declined as the probe frequencies deviated from the target frequency, to about 60% correct for the furthest probes. These results are consistent with findings of other similar experiments (Greenberg & Larkin, 1968; Macmillan & Schwartz, 1975; Scharf et al., 1987; Dai et al., 1991; Schlauch and Hafter, 1991; Tan et al., 2008). The pattern of responses is consistent with participants operating with an attentional filter (or listening band) centred at the target frequency to detect the pure tone stimuli. The skirts of that filter then appear to have an attenuating effect on the level of probe signals as they

deviate from the filter's centre. Importantly, the results of this control experiment demonstrate that the modified probe-signal method utilised in the current experiments are capable of revealing attentional filter mechanisms in listeners.

Given that each of the two T+P conditions contained, at most, two probe frequencies of equal distance from the target; it was possible for participants to employ different listening strategies for each condition. The general pattern, however, was that there was a peak in sensitivity to the target frequency, regardless of whether probes with a  $\Delta f$  of 50 Hz or 100 Hz were presented. In the condition where probes had a  $\Delta f$  of 100 Hz, the participants may have monitored a wider listening band than that for the condition containing probes with a  $\Delta f$  of 50 Hz. If this were the case, however, it might be expected that the performance on the target signal would have been reduced in this condition, with respect to the performance on the target in the condition with a  $\Delta f$  of 50 Hz. This would occur as less attentional resources would be available to allocate to the each frequency region when attending to a wider listening band. Examination of the data, however, revealed no differences between target detection on the T+P condition with a  $\Delta f$  of 100 Hz compared to target detection in the condition with a  $\Delta f$  of 50 Hz. A listening strategy which could better explain the results is if participants monitored the same width listening band across conditions. This could explain why the 100 Hz probes were not detected as well as 50 Hz probes. It could be that the 100 Hz probes were outside the participants' monitored listening bands, whereas probes with a  $\Delta f$  of 50 Hz were within this listening band and thus were identified with greater accuracy than more distant probes.

In order to check whether the presence of probes affects the detection of target signals, responses to targets in T+P conditions were compared to responses for the same signal in the T-only condition. In agreement with the results of Greenberg and Larkin (1968), these comparisons indicated that the presence of probe signals did not influence the

detectability of target signals. The detection rates of the 1 kHz target tones were almost identical in the T+P conditions as they were in the T-only condition.

An important cautionary note for considering the mean results is that there was a large amount of variability across the participants' responses. Additionally, the results did not give such a clear demonstration of each individual's filter characteristics as has been found in previous studies with more highly-practiced participants (Greenberg & Larkin, 1968; Macmillan & Schwartz, 1975). Given that all data were obtained within one experimental session, there was less time available to yield a detailed approximation of the attentional filter of each listener. Unlike the classic experiments of Greenberg and Larkin (1968), the current experiment was comparatively short, sampling the frequency range more sparsely and presenting a much narrower range of frequencies.

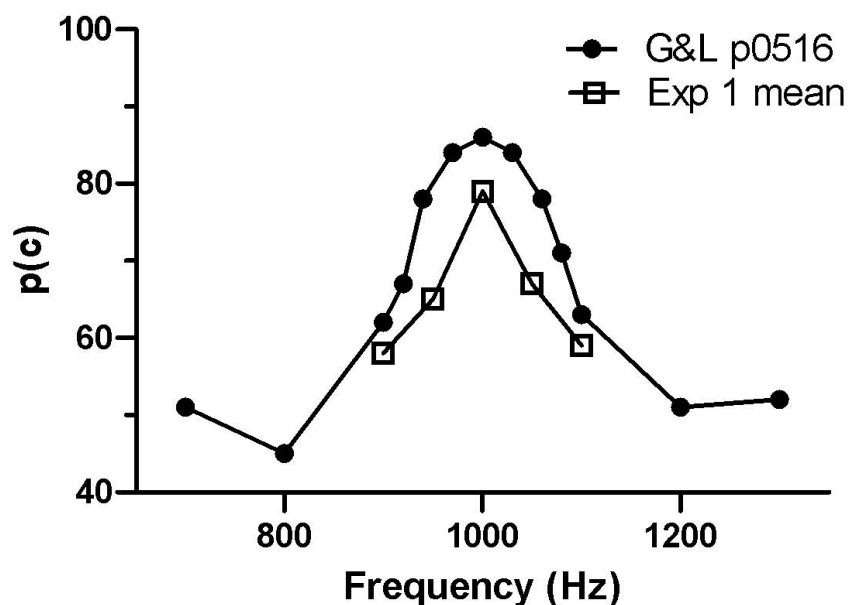
Greenberg and Larkin (1968) suggested that six sessions are required in order to get a good approximation of the response characteristic of any given listener. In their first experiments, all participants completed 24 sessions, each made up of ten blocks of 100 or 94 trials. Their final experiment was reduced to six sessions which they found to be adequate to obtain the necessary data. Similar research making use of the probe-signal method has concluded that 960 trials (which can be completed within a two hour period) are required to get a relatively consistent and stable picture of a subject's attentional filter at one frequency (Dai et al., 1991). In other research, participants received at least seven hours of practice before they began data collection (Macmillan & Schwartz, 1975). Later research, on the other hand, emphasized that little practice is required in order to get adequate estimations of the attentional filter (Scharf et al., 1987). For Experiment 1, previous studies were used to estimate the minimal number of probe frequencies and trials required in order to give a basic estimate of the attentional filter for frequency. Thus, the experimental design consisted of a single session with five blocks of 168 trials, which was completed within an average time

span of one hour. Additionally, the practice session was limited to a small number of 12-trial blocks. Recall that this was a control experiment for the main experiment (experiment 2), and so less time was dedicated to this portion of the study compared to previous studies investigating attentional filters only in the frequency domain. Nevertheless, the design was still effective for obtaining group data that were comparable to those of previous experiments.

Although the individual results are not as detailed as those obtained in previous experiments, all observers tended to show peaks in sensitivity to the target signal. Moreover, for some individuals, a sharp, filter-like function could be measured in this experimental session. Other individuals, however, failed to exhibit this pattern of results. The responses for each individual were different regarding their sensitivity to probe signals. While some individuals tended to show increased sensitivity to higher frequency probes, others tended to show that for lower frequency probes. The likely reason for the high levels of variability across participants is that each individual needed to complete more trials in order for an accurate measure of each person's attentional filter to be obtained. Listeners did not have an adequate number of trials to finely tune their listening to the frequency region of the target. Swets and Kristofferson (1970) state that detection will suffer when the listener is "not tuned as precisely as possible to the frequency that is presented" (p. 350). If this were the case, it is plausible that having less testing time in order to encourage participants to tune their listening to the target frequency should decrease the likelihood that they can detect that tone.

For the furthest probes in this experiment, mean detection rates hovered slightly above chance levels at 60%. Recall that in Greenberg and Larkin's (1968) study, participants' performance declined to chance levels for the most distant probes. When assessing the bandwidth of both sets of results, however, there are remarkable similarities between the two. A demonstration of the mean results of the current experiment and those of a participant from Greenberg and Larkin's (1968) study are displayed in figure 20. From this figure, it is clear

that the slight discrepancy in the detection of the furthest probes is likely due to the difference in the range of frequencies sampled. Greenberg and Larkin (1968) presented a larger range of probes and found that performance declined to chance levels at 300 Hz either above or below the target frequency. The results indicate that the bandwidth is likely to be approximately 400 Hz. The current study did not extend probes beyond 100 Hz from the target. It is therefore a reasonable assumption, that if an additional condition were added to the current experiment, using probes with  $\Delta f$ s of up to 200 Hz, then the performance would decline and indicate a similar bandwidth as that estimated from the participant in Greenberg and Larkin's (1968) study. Another slight difference between the two sets of results presented in Figure 20 is that the overall detection rates in the current experiment were slightly lower than that obtained by Greenberg and Larkin's (1968) participant. This difference is likely to be attributed to the SNRs chosen rather than any difference in the listening strategies of the participants in each experiment.



**Figure 20.** Comparison of mean results of Experiment 1 and results of participant 0516 from the Greenberg and Larkin (1968) experiment

Despite the drawbacks of testing a small frequency range and having fewer trials per condition, the design of this experiment had some advantages over previous research. Firstly, the current research focused on obtaining a larger sample of participants. Previous studies used as few as four participants (Haft et al., 1993; Greenberg & Larkin, 1968) in multiple sessions to derive average listening bands. Having a larger sample meant that each listener could participate in the research in a single session. As a consequence of the significantly shorter experimental time, the obtained results should reflect participant performance which is not hindered by a decline in interest as a reaction to the absolute monotony of the task. Greenberg and Larkin (1968) experienced this problem in their first experiment. A gradual decline in signal detection was observed over the 24 sessions. As a result the experimenters introduced “novelty” blocks into the experiment which involved presenting only target signals at an atypically low intensity level. This was successful for sustaining performance throughout the experiment for three of the four participants. In the current experiment, patient fatigue was not an issue, and so no additional changes to the listening conditions were required in order for participants to maintain concentration on the task.

#### **4.1.2 Selective Attention**

The ability to detect 900 Hz and 1100 Hz signals when they were presented alone was markedly better than the ability to detect these same signals when they were presented at the same SNR as probes. This result is a replication of an experiment which has been more rigorously done several times by multiple experimenters, and illustrates the role of auditory selective attention in this tone-detection task. If participants did not make use of some form of attentional mechanism to respond to the tonal stimuli then they should respond to probe signals in a similar manner to when they are presented as targets. Given that performance for the probe tones with a  $\Delta f$  of 100 Hz was markedly better when they were presented in T-only conditions as targets, as compared to when they were presented at the same level in T+P

conditions as probes – this suggests that attention must play some role in the decision-making process. The results of this experiment serve as a control for the second experiment, by demonstrating that there is an attentional filter mechanism for frequency which can be measured using this modified probe-signal method.

#### **4.1.3 Individual Differences**

The large amount of individual variability may be due to the tendency for listeners to employ different response strategies. A strategy might be defined by the width or the number of bands which a listener monitors during these tasks. At a pre-conscious level, different participants may have assigned more attentional resources to different frequency regions. While some seemed to allocate greater attention to higher frequency regions, others did so for lower frequency regions. The reason underlying these seemingly different patterns of responding remains unclear as a result of the high levels of variability obtained in this experiment.

Veniar (1958) described the possibility that there may be a distribution of listeners who range from narrow-band to broad-band listeners. She also postulated that the most common listener might be one who is able to readily shift from a narrow-band to a broad-band listening strategy for different tasks. It is not possible to tell whether the response patterns of each participant obtained in this study reflect those which were applied on all trials, or whether they represent averages of several different listening strategies, each used on different subsets of trials within the conditions.

Sequential analysis of the data obtained for each participant might uncover changes in listening strategies. Swets, Shipley, McKey and Green (1959) conducted an uncertain-frequency task and found that participants were better able to detect a target signal on a trial if it followed either a correctly detected target or an incorrectly detected probe on the previous trial. This result could be interpreted as being a demonstration of a “positive

recency” effect (Macmillan & Schwartz, 1975) in the setting of their attentional filter, or alternatively as a result of different weights assigned to multiple filters. Further research could look at sequential data to confirm how listening strategies might be altered throughout a probe-signal experiment to attend to either high or low-frequency probes, or narrow or wide frequency bands.

#### **4.1.4 Conclusion**

In summary, the attention function of the participants shows correct detection of the target tone on the majority of trials, and systematically declining toward chance levels for probe frequencies 100 Hz distance at either side of the target frequency. Though there was a large amount of variability in response patterns, the mean frequency response curve shows a distinct similarity to those of previous experiments, and to the frequency-response characteristics of a band-pass filter. The data therefore support the hypothesis that there is a selective attentional filter for frequency. The results from this control experiment can be used to support the methods of the second experiment, which are considered adequate for demonstrating an attentional filter for  $f_0$ , should it exist. The results of Experiment 2 are discussed next.



## **4.2 Experiment 2**

The purpose of experiment 2 was to investigate whether there is an attentional filter in the f0 domain. If there were an attentional filter for f0, its properties may be similar to the attentional filter for frequency. The modified probe-signal method used in experiment 1 was successful in revealing an attentional filter for frequency. Therefore, the same method was used, replacing pure tone stimuli with sentence stimuli, in order to investigate whether a similar attentional filter can be found for f0. If such a filter did exist, when attention is drawn toward sentences with a certain f0 (target), it is expected that the sensitivity to sentences with differing f0s (probes) would be reduced.

### **4.2.1 Experimental Findings**

The major outcome of experiment 2 was that the attention functions relating to the identification of sentence words with target and probe f0s did not show the filter-like characteristics which were demonstrated in experiment 1 using pure tone stimuli. In other words, the selectivity provided by an attentional filter for frequency was not replicated in the f0 domain. These results can be interpreted in two ways. It could be that this form of selective attention is not preserved for speech stimuli. Alternatively, it could be that the attentional filter characteristics for f0 are such that a more detailed experiment is required to measure how the filter operates. In any case, the differences between the results of the two experiments cannot be attributed to the methods used to obtain the data.

Although the mean attention function does not look like that of the attentional filter, there was an interesting pattern of results which was not anticipated. There was a difference in the mean rate of correctly identified sentences between those with the target f0 and those with a  $\Delta f0$  minus 5 Hz. The target sentences were identified significantly better than probe sentences with a  $\Delta f0$  of -5 Hz. A similar trend was found for the probe sentences with a  $\Delta f0$

of +5 Hz; however there was no significant difference when comparing these results against those for the target. In spite of these findings, the identification rates for 10 Hz probe sentences were about the same as those for target sentences.

Similar to experiment 1, these mean results must be interpreted with caution. There was a large range of response patterns across the participants. While some participants seemed to clearly demonstrate poorer identification of probe sentences with a  $\Delta f_0$  of 5 Hz as compared to target sentences and probe sentences with a  $\Delta f_0$  of 10 Hz, others showed no differences in their responses to the different  $f_0$  sentences.

## **4.2.2 Attention**

### **4.2.2.1 The possibility of $f_0$ cueing**

It is possible that there was cueing of the  $f_0$  for each sentence before the to-be-identified colour-number pairs were presented. In experiments with pure tones, when auditory cues are presented before the to-be-detected signal, detection rates improve if the cue matches the target signal in frequency (Tan et al., 2008). Recall the sentence structure: “GO TO ‘COLOUR’ ‘NUMBER’ NOW” used in the current experiment. All words within the sentence were manipulated to have the same  $f_0$  as the words to be identified. It could be that the non-target words “GO TO” were enough to cue the participant to change which  $f_0$  region they should attend to. If this were the case, on every trial, the participant would selectively change which  $f_0$  they were attending to, in time to use this as a cue for identifying the colour-number pair at the end of each sentence. This interpretation is consistent with the attention functions of participants who did not show any difference in sentence identification as a function of differences in  $f_0$ .

The above interpretation can only partially account for the overall pattern in attention functions, however. Although cueing for  $f_0$  could explain why performance on probe sentences with a  $\Delta f_0$  of 10 Hz was equal to that of the target sentences, it does not explain the

reduced performance for probes with a  $\Delta f_0$  of 5 Hz. If low-level attention was directed toward the  $f_0$  of each sentence, and was capable of being switched at the beginning of each carrier phrase; this switching should occur for all probe  $f_0$ s, not only the probes with a  $\Delta f_0$  of 10 Hz. The average response for the probes with a  $\Delta f_0$  of 5 Hz was what might be expected if there were an attentional filter for  $f_0$  (i.e., reduced with respect to the target). Therefore, the presence of  $f_0$  filters cannot be ruled out. It might be interesting to investigate the effect of flattening the  $f_0$  of only the colour-number pair, while leaving the rest of the phrase unchanged. This would remove the possibility of cueing for the  $f_0$  at the beginning of each sentence.

#### **4.2.2.2 Top-down Influences**

One possible reason for the inconsistencies found between attention functions obtained in the frequency dimension and those in the  $f_0$  dimension, is that the latter may be controlled to a greater extent by top-down processes from higher order structures. Given that the participants each have years of experience in listening to speech sounds, they are likely to be more attuned to these sounds. This is in contrast to pure tone stimuli, which participants are unlikely to be exposed to in their daily lives. Experience in listening to speech and complex-sounds may make the processing of those sounds more susceptible to attention. It could be that more conscious, as opposed to pre-conscious, attention was being utilised in the task for experiment 2. If people are consciously attending to the changes in  $f_0$ , they may be more likely to switch attention to different  $f_0$ s as they become more different from that which is expected. This might explain why there was filter-like responding for probes having a  $\Delta f_0$  of 5 Hz, however the performance for probes with a  $\Delta f_0$  of 10 Hz was akin to that of the target. A  $\Delta f_0$  of 5 Hz might be small enough that selective pre-conscious filtering takes place, resulting in reduced identification of that sentence. A  $\Delta f_0$  of 10 Hz, however, might be enough to activate conscious attention, which would then allow the listener to switch to the

unexpected probe signal and track the  $f_0$ , thus making use of the  $f_0$  cue. Unfortunately there is no way to measure the extent to which different types of the attention are being operated in these experiments.

### 4.2.3 Fundamental Frequency

Another possible explanation for not finding a consistent  $f_0$  effect is that there is indeed an internal attentional filter in the  $f_0$  domain which can influence the identification of speech in noise; however this filter is narrower than allowed for here. In the design of this experiment, the fact that listeners are extremely good at detecting differences in  $f_0$  was considered. Thus the  $\Delta f_0$ s chosen for the probes in the current experiment were 5 and 10 Hz. It is entirely possible, however, that any attentional filter which is tuned to  $f_0$ , operates over a range between 10 and 20 Hz, which could explain the poorer identification rate of probe sentences with a  $\Delta f_0$  of 5 Hz but not the probe sentences with a  $\Delta f_0$  of 10 Hz. If this hypothesis were correct, a  $\Delta f_0$ , of say 7 Hz, might result in even poorer identification performances than that which was obtained for the probes with a  $\Delta f_0$  of 5 Hz. Additionally, if smaller  $\Delta f_0$ s (such as 2 Hz) were used, it would be expected that identification would be better than that which was obtained for probes having a  $\Delta f_0$  of 5 Hz. Further research is required in order to investigate these possibilities.

If the probe sentences with a  $\Delta f_0$  of 10 Hz are outside the attentional filter tuned to the target  $f_0$ , previous research using pure tones would predict that identification of these sentences would remain as poor as that which is obtained at the boundaries of the attentional filter. In other words, the attention function is thought to be flat beyond the attentional filter (Dai et al., 1991). This interpretation is inconsistent with the results of the present experiment. The findings show better identification of the probe sentences with a  $\Delta f_0$  of 10 Hz compared to those with a  $\Delta f_0$  of 5 Hz. It is plausible, however, that the presence of any attentional filters for  $f_0$  may differ in their characteristics from those on other dimensions

(such as pure tones). Speech stimuli with  $f_0$ s outside the expected region may be enough to shift a listener's attentional filter from one  $f_0$  to another, resulting in identification rates which parallel those of target or expected  $f_0$  speech. Exactly what  $\Delta f_0$  might be required in order for listeners to shift their attention to another  $f_0$  cannot be ascertained from this experiment.

There are reasons, however, to cast doubt upon the possibility that the current experiment has 'missed' the true shape of the attentional filter for  $f_0$ . Firstly, if the results did resemble a filter-like characteristic response pattern, then the reduced identification rate for the -5 Hz probes should be matched with an equally low identification performance on the +5 Hz probes. Only the -5 Hz  $\Delta f_0$  yielded a significant change in identification. Secondly, although the sentences with a  $\Delta f_0$  of 5 Hz overall were identified at a lesser rate than those for the target and 10 Hz  $\Delta f_0$  sentences, the difference was relatively small (ranging from about 5 to 10%). Additionally, it would seem that while some participants showed this pattern of response, others did not show any differences in their response patterns to the target and probe sentences. Could it be that some people exhibit this tuning pattern and others do not? This is unlikely, given that previous experiments investigating internal filter mechanisms have found these to be relatively fixed across the normal hearing population (Moore, 1987).

#### **4.2.4 Possible Contributing Factors**

There are a number of other factors which may have had a role in the outcome of this experiment. Those which stem from the chosen design include: time restraints and practical concerns with regard to the use of speech stimuli in a probe-signal method of experimentation. The effect of individual variability is also discussed here.

##### **4.2.4.1 The Effect of the Number of Trials and Test Time**

The lack of consistent findings across participants could be attributed to the small number of trials dedicated to each condition for experiment 2. In the current study, emphasis was made on ensuring that the chosen methodology was adequate. In order to do this, participants completed the tone-detection experiment as well as the sentence-identification task. The tone-detection experiment was an important component of this study as it provided evidence that the attentional filter for frequency was measurable with the chosen modified version of the probe-signal method. Unfortunately, as a result of this, there was a limit to the number of trials which could be presented to obtain results for Experiment 2. To minimise testing time and consequent participant fatigue, all participants completed one block of each condition. This small number of trials (total of 504) may not have been adequate to obtain a detailed and reliable measure of each participant's response patterns for more complex stimuli such as those used in this experiment. Future studies which investigate the ability of the effect of  $f_0$  on the ability to identify speech in noise may find more conclusive results if a larger number of trials is implemented. Due to practical considerations, this would need to be completed in multiple experimental sessions.

#### **4.2.4.2 The Effect of Stimulus Factors**

The listening conditions in this experiment were in some ways, optimal, for focusing on the target words of the speech signal. Although the speech was presented in noise, the  $f_0$  was highly predictable and the target words were always presented at the same point for every sentence. This is not very representative of the listener's ability to identify speech in the real world. In reality, even if a person's voice and underlying  $f_0$  is familiar, timing cues for the important segments of a sentence may not be available. These laboratory settings may have given participants an advantage for identifying the target words in each sentence.

On the other hand, after manipulating the chosen talker's sentences to have a flat (controlled)  $f_0$ , prosodic cues such as intonation, volume and tempo changes were removed.

In this respect, the experimental design may have put participants at a disadvantage for identifying the target words in the sentences, as compared to speech identification in a real-world setting. Indeed, there is evidence to suggest that intonation is a valuable cue which assists in speech reception (Brokx & Nootboom, 1982; Binns & Culling, 2007). When a target sentence is played simultaneously with a masking sentence having different f0s, the percentage of errors is approximately 15% higher when intonation contains f0 variations compared to when f0 variations are removed (Brokx & Nootboom, 1982). For the current study, it is not known whether the number of incorrect responses for the colour-number pairs would have changed as a result of having variations in the f0. It may be the case that when speech sounds have different f0s, it is easier to attend to one particular f0 when that f0 is allowed to vary naturally. It was not possible to maintain intonation in the present experiment, since a constant f0 was required to investigate the sole contribution of f0 cues. In summary, the reader should be aware that f0 did not vary naturally, and so the results might not be representative of how attention is allocated to f0 in real-world settings.

A possible downfall of the stimuli used in this experiment is that the language spoken was American English. The participants, on the other hand, were speakers of New Zealand English. Participants may have had trouble identifying the colour-number pairs as a result of unfamiliarity with the accent of the American English speaker. This problem was presumably eliminated when the SNR was chosen for each participant. For example, speakers of American English might have been able to be tested at a lower SNR than the participants in this experiment. It cannot be ruled out, however, that there may have been residual effects of accent which make some sentences more difficult for the participants to identify than others. If this study were to be replicated, it might be useful to use a set of sentences spoken in New Zealand English to remove this possibility.

In an ideal world, the participants would have been listening for a familiar and natural New Zealand talker's voice as the target signal. For the purpose of presenting probe sentences, another talker with only a slight change in  $f_0$  could be used. Unfortunately it is not possible to find two speakers with underlying  $f_0$ s which differ in a controlled manner.

#### **4.2.4.3 The Effect of Individual Differences**

There was a large range of possible response characteristics obtained in this experiment. The variability was greater than that which was measured in experiment 1. Once again, the possibility of participants adopting different listening strategies should be considered. Recall that some participants did not show any difference in their responses to sentences of differing  $f_0$ s, whereas others were better at identifying words with the target  $f_0$  or probe sentences with a  $\Delta f_0$  of 10 Hz than those with  $\Delta f_0$  of 5 Hz. One explanation for this lies in the differences between the complexity of pure tones and speech. Previous researchers have suggested that individual differences in auditory processing typically increase as the stimuli and task become more complex (Neff & Dethlefs, 1995). This is thought to reflect, at least in part, differences between individuals' "rate of perceptual learning" (p. 125) which varies with stimulus and task complexity (Neff & Dethlefs, 1995). This could explain the greater variability in the second experiment. Compared to pure tones, the speech stimuli are more complex. This may have led to more pronounced differences between the ways in which participants allocate attentional resources. Therefore, different individuals might make use of  $f_0$  in different ways which cannot be measured directly in psychophysical experiments.

#### **4.2.5 Conclusions and Directions for Future Research**

The present application of the probe-signal method to  $f_0$  for speech stimuli demonstrates that the listeners' attention band does not approximate the shape of the attentional filter, obtained in typical pure tone experiments. Overall results yielded a pattern



which approximated a filter-like function for probes with a  $\Delta f_0$  of 5 Hz; however this pattern did not extend to probes with a  $\Delta f_0$  of 10 Hz. Perhaps an attentional filter for  $f_0$  is narrower than tested for in this experiment, reflecting heightened sensitivity to these changes in speech. This would make sense for listening in the real world, given that it is often necessary to notice these changes.

In order to say more conclusively whether or not an attentional filter mechanism might exist for  $f_0$ , further research is encouraged. Future research on this topic should aim to use smaller and larger  $\Delta f_0$ s (e.g. 2 and 20 Hz) than was used in the current experiment. Furthermore, a more densely sampled  $f_0$  range is required. These strategies would enable researchers to explore the possibilities both of an attentional filter which operates over an extremely narrow band of  $f_0$ s, and also to find the  $\Delta f_0$  required for attention to be switched to a different  $f_0$  to assist in speech identification.

A deeper understanding of the way in which attention is allocated to  $f_0$  in the presence of background noise is required in order to understand how  $f_0$  can be used as an acoustic cue for speech understanding. This might assist in the development and accuracy of  $f_0$  extraction algorithms in digital signal processing for hearing instruments. As far as we are aware, this is the first experiment of this type to have been attempted. Further research is therefore needed to determine if an attentional filter in the  $f_0$  domain is present or not, and moreover the shape and width of this, should it exist.

### **4.3 *Summary and Conclusions***

Twenty participants served in two experiments using the probe-signal method. This method allows us to measure the characteristics of attentional filters for different acoustic dimensions, should they exist. Attention functions were measured for frequency and  $f_0$  in these experiments. The response curves obtained reveal the range of frequencies or  $f_0$ s to which the participants were likely to be attending to. The first experiment replicated findings

of past research indicating that listeners demonstrate differential responding as a function of the frequency of the signal. These results support a filter model of attention in the frequency domain. The results of the second experiment contained greater variability and the data do not directly support a filter-like attention function in the  $f_0$  domain. Given that the two experiments aimed at measuring different acoustic dimensions, the differences in responding can be interpreted in various ways. Future research should look to replicate research using the probe-signal method in the  $f_0$  domain, using a wider and more densely sampled  $f_0$  range in multiple sessions to get a more reliable estimate of the influence of low-level attention and  $f_0$  for speech identification.

## References

---

- Arehart, K. H. (1998). Effects of high-frequency amplification on double-vowel identification in listeners with hearing loss. *The Journal of the Acoustical Society of America*, 104(3), 1733–1736.
- Arehart, K. H., King, C. A., and McLean-Mudgett, K. S. (1997). Role of fundamental frequency differences in the perceptual separation of competing vowel sounds by listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research*, 40(6), 1434–44.
- Arehart, K. H., Rossi-Katz, J., & Swensson-Prutsman, J. (2005). Double-vowel perception in listeners with cochlear hearing loss: Differences in fundamental frequency, ear of presentation, and relative amplitude. *Journal of Speech, Language, and Hearing Research*, 48(1), 236–252.
- Assmann, P. F. & Summerfield, Q. (1990). Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *The Journal of the Acoustical Society of America*, 88(2), 680–697.
- Assmann, P. F., & Summerfield, Q. (1994). The contribution of waveform interactions to the perception of concurrent vowels. *The Journal of the Acoustical Society of America*, 95(1), 471–484.
- Atkinson, J. E. (1978). Correlation analysis of the physiological factors controlling fundamental voice frequency. *Journal of the Acoustical Society of America*, 63(1), 211–22.
- Baker, R. J., & Rosen, S. (2006). Auditory filter nonlinearity across frequency using simultaneous notched-noise masking. *The Journal of the Acoustical Society of America*, 119(1), 454–462.
- Bernstein, J. G. W., & Oxenham, A. J. (2006). The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level. *The Journal of the Acoustical Society of America*, 120(6), 3916–3928.
- Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception & Psychophysics*, 64(5), 844–854.
- Binns C., & Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *Journal of the Acoustical Society of America*, 122, 1765.
- Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America* 107, 1065–1066.

- Broadbent, D. E., & Ladefoged, P. (1957). On the fusion of sounds reaching different sense organs. *Journal of Acoustical Society of America*, 29(6), 708-710.
- Broadbent, D. E. (1958). *Perception and Communication*. New York: Pergamon Press.
- Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, 10(1), 23-36.
- Brungart, D. S. (2001). Evaluation of speech intelligibility with the coordinate response measure. *The Journal of the Acoustical Society of America*, 109(5), 2276-2279.
- Carroll, J., Tiaden, S., & Zeng, F. (2011). Fundamental frequency is critical to speech perception in noise in combined acoustic and electric hearing. *The Journal of the Acoustical Society of America*, 130(4), 2054-2062.
- Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, 25(5), 975-979.
- Cruttenden, A. (1986). *Intonation*. Cambridge [Cambridgeshire]: Cambridge University Press.
- Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology. Human Perception and Performance*, 30(4), 643-656.
- Cutting, J. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review*, 83(2), 114-140.
- Da Costa, S., van der Zwaag, W., Miller, L. M., Clarke, S., & Saenz, M. (2013). Tuning in to sound: Frequency-selective attentional filter in human primary auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(5), 1858-1863.
- Dai, H., Scharf, B., & Buus, S. (1991). Effective attenuation of signals in noise under focused attention. *The Journal of the Acoustical Society of America*, 89(6), 2837-2842.
- Dai, H., & Wright, B. A. (1995). Detecting signals of unexpected or uncertain durations. *The Journal of the Acoustical Society of America*, 98(2), 798-806.
- Deutsch, J. A., & Deutsch, D. (1963) Attention: Some theoretical discussions. *Psychological Review*, 70(1), 80-90.
- Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, 92(Pt 1), 53-78.
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, 87(3), 272-300.

- Eddins, D. A., & Liu, C. (2012). Psychometric properties of the coordinate response measure corpus with various types of background interference. *The Journal of the Acoustical Society of America*, 131(2), 177-183.
- Fletcher, H. (1940). Auditory Patterns. *Reviews of Modern Physics*, 12(1), 47-65.
- Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data, *Hearing Research*, 47(1), 103–138.
- Green, D. M. (1958). Detection of multiple component signals in noise. *The Journal of the Acoustical Society of America*, 30(10), 904-911.
- Green, D. M. (1961). Detection of auditory sinusoids of uncertain frequency. *The Journal of the Acoustical Society of America*, 33(7), 897-903.
- Green, T., Faulkner, A., Rosen, S., & Macherey, O. (2005). Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification, *The Journal of the Acoustical Society of America*, 118(1), 375–385.
- Greenberg, G. Z., & Larkin, W. D. (1968). Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: the Probe-Signal Method. *The Journal of the Acoustical Society of America*, 44(6), 1513-1523.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Haftner, E., Sarampalis, A., & Loui, P. (2007). Auditory Attention and Filters. In W. Yost, A. Popper & R. Fay (Eds.), *Auditory Perception of Sound Sources* (Vol. 29, pp. 115-142). United States: Springer US.
- Haftner, E. R., Schlauch, R. S., & Tang, J. (1993). Attending to auditory filters that were not stimulated directly. *Journal of the Acoustical Society of America*, 94(2), 743-747.
- Haukia, M. H., & Bregman, A. S. (1984a). Perceptual segregation of simultaneous vowels presented as steady states and as parallel and crossing glides. *Journal of the Acoustical Society of America*, 75(S1), 83.
- Haukia, M. H., & Bregman, A. S. (1984b). Perceptual segregation of simultaneous vowels presented as steady states and as glides. *Canadian Psychology*, 25, 210.
- Hübner, R., & Haftner, E. R. (1995). Cuing mechanisms in auditory signal detection. *Perception & Psychophysics*, 57(2), 197.
- Ison, J. R., Virag, T. M., Allen, P. D., & Hammond, G. R. (2002). The attention filter for tones in noise has the same shape and effective bandwidth in the elderly as it has in young listeners. *The Journal of the Acoustical Society of America*, 112(1), 238 - 246.
- Macmillan, N. A., & Schwartz, M. (1975). A probe-signal investigation of uncertain-frequency detection. *Journal of the Acoustical Society of America*, 58(8), 1051 – 1058.

- McDermott, H. J., & McKay, C. M. (1994). Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea. *Journal of the Acoustical Society of America*, 96(1), 155–162.
- Mondor, T. A., & Bregman, A. S. (1994). Allocating attention to frequency regions. *Perception & Psychophysics*, 56(3), 268-276.
- Moore, B. C., & Glasberg, B. R. (1983). Suggested formula for calculating auditory filter bandwidth and excitation patterns. *Journal of the Acoustical Society of America*, 74, 750-753.
- Moore, B. C. J. (1987). Distribution of auditory-filter bandwidths at 2 kHz in young normal listeners, *Journal of the Acoustical Society of America*, 81, 1633–1635.
- Moore, B. C., Hafter, E. R., & Glasberg, B. R. (1996). The probe-signal method and auditory-filter shape: Results from normal- and hearing-impaired subjects. *The Journal of the Acoustical Society of America*, 99(1), 542 - 552.
- Moore, B. J., & Carlyon, R. (2005). Perception of Pitch by People with Cochlear Hearing Loss and by Cochlear Implant Users. In C. Plack, R. Fay, A. Oxenham & A. Popper (Eds.), *Pitch* (Vol. 24, pp. 234-277): Springer New York.
- Moore, B. C. J., & Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *The Journal of the Acoustical Society of America*, 91(5), 2881-2893.
- Moore, B. C. J., Peters, R. W., & Glasberg, B. R. (1990). Auditory filter shapes at low center frequencies. *The Journal of the Acoustical Society of America*, 88(1), 132 – 140.
- Moore, B. C. J. (2012). *An introduction to the psychology of hearing*. Bingley: Emerald.
- Moray, N. (1960). Broadbent's filter theory – postulate H and the problem of switching time. *Quarterly Journal of Experimental Psychology*, 12, 21-220.
- Moray, N. (1969). *Attention: Selective processes in vision and hearing*. London: Hutchinson Educational.
- Neff, D. L., & Dethlefs, T. M. (1995). Individual differences in simultaneous masking with random-frequency, multicomponent maskers. *The Journal of the Acoustical Society of America*, 98(1), 125.
- Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America*, 59(3), 640-659.
- Patterson, R. D., & Moore, B. C. J. (1986). Auditory filters and excitation patterns as representations of frequency resolution. In B. C. J. Moore (Ed.), *Frequency selectivity in hearing* (pp. 123 – 177). London: Academic Press.

- Penner, M. J. (1972). The effect of payoffs and cue tones on detection of sinusoids of uncertain frequency. *Perception & Psychophysics*, 11(3), 198-202.
- Reeves, A. (2013). The auditory attention band. In C. Chubb, B. A. Doshier, Z. Lu, R. M. Shiffrin (Eds.) , *Human information processing: Vision, memory, and attention* (pp. 193-205). Washington, DC: American Psychological Association.
- Rogers, W. L. & Bregman, A. S. (1993). An experimental evaluation of three theories of auditory stream segregation. *Perception & Psychophysics*, 53, 179-189.
- Scharf, B., Magnan, J., & Chays, A. (1997). On the role of the olivocochlear bundle in hearing: 16 case studies. *Hearing Research*, 103(1-2), 101-122.
- Scharf, B., Quigley, S., Aoki, C., Peachey, N., & Reeves, A. (1987). Focused auditory attention and frequency selectivity. *Perception & Psychophysics*, 42(3), 215-223.
- Schlauch, R. S., & Hafter, E. R. (1991). Listening bandwidths and frequency uncertainty in pure-tone signal detection. *The Journal of the Acoustic Society of America*, 90(3), 1332-1339.
- Schouten, J. F. (1940). The residue and the mechanism of hearing. *Proceedings of the Koninklijke Akademie van Wetenschap*, 43, 991-999.
- Simon, H. J., & Yund, E. W. (1993). Frequency discrimination in listeners with sensorineural hearing loss. *Ear and Hearing*, 14(3), 190-201.
- Speith, Curtis & Webster (1954). Responding to one of two simultaneous messages. *Journal of the Acoustic Society of America*, 26(3), 391-396.
- Summers, V., & Leek, M. R. (1998). FO processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *Journal of Speech, Language, and Hearing Research*, 41(6), 1294 - 1306.
- Swets, J. A., Shipley, E.G., McKey, M. J., & Green, D. (1959). Multiple observations of signals in noise. *The Journal of the Acoustical Society of America*, 31(4), 514-521.
- Swets, J. A. (1963). Central factors in auditory frequency selectivity. *Psychological Bulletin*, 60(5), 429-440.
- Swets, J. A., & Kristofferson (1970). Attention. *Annual Review of Psychology*, 21(1), 339-366.
- Tanner, W. P., Jr., & Norman, G. Z. (1954). The human use of information. II. Signal detection for the case of an unknown signal parameter. *Transactions on Professional Group on Information Theory, PGIT-4*, 222-226.
- Tanner, W. P., Jr. (1956). Theory of Recognition. *Journal of the Acoustical Society of America*, 28, 882-888.

- Tanner, W. P., J., Swets, J. A., & Green, D. M. (1956). Some general properties of the hearing mechanism. In: Swets, J. A. (1963). Central factors in auditory frequency selectivity. *Psychological Bulletin*, 60(5), 429-440.
- Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12(4), 242—248.
- Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, 3(6), 449–201.
- Van Noorden, L. P. A. S. (1975). Temporal coherence in the perception of tone sequences. PhD. Thesis, Eindhoven University of Technology, Eindhoven.
- Veniar, F. A. (1958). Signal detection as a function of frequency ensemble. I. *The Journal of the Acoustical Society of America*, 30(11), 1020-1024.
- Wright, B. A., & Dai, H. (1998). Detection of sinusoidal amplitude modulation at unexpected rates. *The Journal of the Acoustical Society of America*, 104(5), 2991-2996.
- Yost, W. A., Sheft, S. S. (1993). Auditory perception. In: W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Human Psychophysics*. (pp. 193 – 236). New York: SpringerVerlag.
- Zhou, B. (1995). Auditory filter shapes at high frequencies. *The Journal of the Acoustical Society of America*, 98(4), 1935-1942.
- Zwicker, E., & Terhardt, E. (1982). Analytical expression for critical band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America*, 68(5), 1523-1525.



\*\* Licensing permission to reproduce figures was kindly granted by the copyright holders of each figure cited throughout this work.

# Appendix A. Participant Information Sheet

---

## UNIVERSITY OF CANTERBURY HUMAN ETHICS COMMITTEE APPLICATION FOR REVIEW & APPROVAL



Department of Communication Disorders

### Project Information Sheet

#### PARTICIPANT INFORMATION

You are invited to participate in the research project entitled *The Effects of Fundamental Frequency and Auditory Selective Attention on the Perception of Speech in Noise*

The aim of this project is to evaluate the effects of attention on understanding speech in background noise. Your involvement in this project will involve one session, lasting approximately 1 and a half hours including rest breaks as needed, during which you will listen to sounds presented over headphones in background noise. The first experiment requires you to listen to a set of sounds where the task involves detecting tones in the presence of background noise. In the second experiment, you will be asked to listen to various sentences in the presence of background noise. After each sentence presentation you will identify two key words that you heard by clicking on a box on a visual display. This task requires a relatively high level of concentration over a prolonged testing period, and so participants are encouraged to take breaks as needed throughout the experiment.

You have the right to withdraw from the project at any time for any reason without penalty, including withdrawal of any information provided or data collected. Withdrawal will not affect any ongoing or future relationship with the University of Canterbury Speech and Hearing Clinic or the Department of Communication Disorders.

The results of the project may be published, and a Master's Thesis is a public document, accessible via the University of Canterbury library database but you may be assured of the complete confidentiality of data gathered in this investigation: the identity of participants will not be made public. To ensure anonymity and confidentiality, the information gathered will be assigned a number and all identifiable information removed. Data and back-up files will be kept on hard drives which are accessible only to the investigators. This data will be kept for ten years after which time it will be destroyed.

The project is being carried out as a requirement for a Masters of Audiology by Rachel Peddie under the supervision of Donal Sinex. The project has been reviewed and approved by the University of Canterbury Human Ethics Committee. If you have any further questions about the research project, please do not hesitate to contact either my supervisor or myself at the University of Canterbury. Thank you once again.

Sincerely,

Rachel Peddie  
Master of Audiology Student

Mob: 027 334 8409  
Email: rnp37@uclive.ac.nz

Donal Sinex  
Dept of Communication Disorders

Ph: 364 2987 extn 7851  
Email: donal.sinex@canterbury.ac.nz

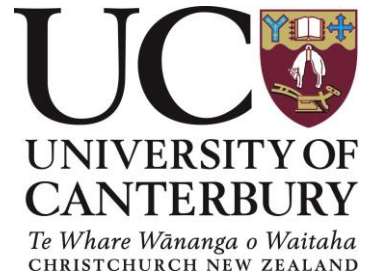
Dept of Communication Disorders, University of Canterbury Private Bag 4800, Christchurch 8140, New Zealand.  
Tel: +64 3 364-2987 Extn 7077  
[www.cmds.canterbury.ac.nz](http://www.cmds.canterbury.ac.nz)

Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand  
Telephone +64 3 364 2987 Extn 45588  
[human-ethics@canterbury.ac.nz](mailto:human-ethics@canterbury.ac.nz)

## Appendix B. Participant Consent Form

---

Department of Communication Disorders



*Rachel Peddie*  
*Department of Communication Disorders*  
*University of Canterbury*  
**Private Bag 4800**  
*Christchurch*

### Consent Form

*The Effects of Fundamental Frequency and Auditory Selective Attention on the Perception of Speech in Noise*

I have read and understood the description of the above-named project. On this basis, I agree to participate in the project, and I consent to publication of the results of the project with the understanding that my anonymity will be preserved.

I understand also that I may withdraw from the project at any time or for any reason, without penalty. I understand that withdrawal will not affect any ongoing or future relationship with the University of Canterbury Speech and Hearing Clinic or the Department of Communication Disorders.

NAME (please print): .....

Signature:

Date:

Dept of Communication Disorders, University of Canterbury Private Bag 4800, Christchurch 8140, New Zealand.  
Tel: +64 3 364-2987 Extn 7077  
[www.cmds.canterbury.ac.nz](http://www.cmds.canterbury.ac.nz)

Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand  
Telephone +64 3 364 2987 Extn 45588  
[human-ethics@canterbury.ac.nz](mailto:human-ethics@canterbury.ac.nz)

## Appendix C. Instructions to Participants

---

### *Experiment 1: Pure tone detection*

- You will hear two intervals of sounds. One will contain noise alone and one will contain noise and a tone
- Your task is to identify the interval that contains a tone plus noise. Do this by clicking on one of the two intervals on the computer screen
- The tone may change in frequency throughout the experiment
- You have unlimited response time on each trial
- You will receive feedback on every trial as the correct interval will flash white after you make your response
- You will have a practice session before the experiment to make sure you know what you are listening for and how to perform the task
- This practice session will be used to find a signal level that will allow you to detect between 60 and 90% of the tones correctly. This signal level will be used for the experimental conditions that follow
- Each experimental condition will last approximately 7 – 8 minutes depending on your response times
- Breaks will be offered every 25 trials. The experiment will pause and ask you to “Click to start/resume”
- Each condition contains 168 trials. At the end of each condition a window will come up asking you if you want to save the data. At this point please raise your hand and I will come through to the booth and load the next condition for you

### *Experiment 2: Sentence identification*

- You will hear a female talker saying a sentence in the presence of background noise
- Your task is to identify the colour-number pair that you hear in the sentence by clicking on the corresponding colour-number pair on a grid on the computer screen
- You have unlimited response time on each trial
- You will receive feedback on every trial as the correct colour-number pair will flash a light blue colour after you make your response
- You will have another practice session before the experiment. This will be used to find a signal level which allows you to identify between 40 and 60% of the sentences correctly. This signal level will be used for the experimental conditions that follow.

## Appendix D. Summary of Speech Manipulation

---

- Praat speech analysis software (Phonetic Sciences, Amsterdam) was used to manipulate each sentence
- Each utterance was read from a file, and selected to create a manipulation object.
- This object was selected for editing.
- In the manipulation window, the pitch contour was stylised to reduce the number of manipulation points in each sentence. The remaining pitch points were then dragged to set the desired  $f_0$  for each point.
- Once the stylised pitch contour was flattened for each sentence, a new sound object was created by selecting the “Get resynthesis (overlap-add)” function.
- The final step in the manipulation was to write the object as a “WAV” file, which was saved in appropriate folders for retrieval.